Complex Word-Formation and the Morphology-Syntax Interface

Susanna Padrosa Trias

PhD Dissertation

Supervisors: Dr Anna Bartra Kaufmann

Dr Jaume Mateu Fontanals

Tutor: Dr Montserrat Capdevila Batet

Departament de Filologia Anglesa i de Germanística

Facultat de Filosofia i Lletres

Universitat Autònoma de Barcelona

May 2010

To my parents, Josep and Mª Dolors,
To my husband, David, and to our son, Dídac,
For all they mean to me.

Table of Contents

Acknowledgments

It is a pleasure to have the opportunity to thank the many people who have contributed to this dissertation in some way, although inevitably some names might be left out. My first debt is to my supervisors, Anna Bartra and Jaume Mateu, who have been a source of knowledge and inspiration, a model to follow, and the perfect friends in a long journey. I also thank them for meeting deadlines, providing me with relevant references, allowing me to pursue my own ideas, sometimes divergent from theirs, and for trusting in me during all these years until the dissertation has reached an end.

I have also benefited from two research stays, the first of which was at University College London, where the topic of this dissertation was decided, mainly due to Ad Neeleman's enthusiasm for morphology. I thank Ad for discussing parts of this dissertation. My second research stay was at the University of Edinburgh, where I owe special thanks to Peter Ackema - who hosted me in the Department of Linguistics and English Language during my stay - for his hospitality. I am also grateful to Peter for providing me with generous comments at various stages in my work.

Much of the material of this dissertation has been presented in conferences like the *Colloquium on Generative Grammar* 16, 17 and 19, held in Madrid, Girona and Vitoria respectively; *Forum de Morphologie / « Décembrettes »*, held in Toulouse; and *Incontro di Grammatica Generativa* 33, held in Bologna. I would like to thank the audiences for comments and suggestions. Some names can be singled out: Víctor Acedo, Peter Ackema, Antonietta Bisetto, Geert Booij, Hagit Borer, Carmen Scherer, Antonio Fábregas, Susana Huidobro, Chiara Melloni, Fabio Montermini, Ad Neeleman, Paco Ordoñez , and Sergio Scalise. I would also like to thank the audiences from the LyCC Colloquium Series in Madrid (CSIC). Special thanks to Violeta Demonte, Olga Fernández Soriano, Nino Grillo, and Carlos Piera. I have also benefited from the seminars organised by the Centre for Theoretical Linguistics at UAB and the members who constitute them: Eulàlia Bonet, Josep Maria Brucart, Teresa Cabré, Verónica Castillo, M. Teresa Espinal, Elías Gallardo, Ángel J. Gallego, Gemma Gómez, Yurena Gutiérrez, M. Lluïsa Hernanz, Wojtek Lewandowski, Angelina Markova, Núria Martí, Silvia Martínez, Joan Mascaró, Ía Navarro, Carme Picallo, Pilar Prieto, Cristina Real, Gemma Rigau, Ester Sánchez, Xico Torres, Maria del Mar Vanrell, and Xavier Villalba.

For various reasons, I am also indebted to Artemis Alexiadou, Mark Baker, Reka Benczes, Reineke Bok-Bennema, Cristina Buenafuentes, Gretel De Cuyper, Anne

Dagnac, Heidi Harley, Richard Kayne, Brenda Laca, Beth Levin, Alec Marantz, Makiko Mukai, Amelia Rus, Peter Svenonius, and Mireille Tremblay. Thanks also to Catherine Dickie, Elspeth Edelstein, Jonathan MacDonald and Andrew Woodard for providing me with native speakers' judgments.

From the Department de Filologia Anglesa i de Germanística at the Universitat Autònoma de Barcelona, I am grateful to Montserrat Capdevila (my tutor), Hortènsia Curell, Michael Kennedy, Mireia Llinàs, Alan D. Reeves, and Maria Josep Solé, all of them being points of reference in my undergraduate studies.

I also want to thank Soledad Varela, Mireia Llinàs, Peter Ackema, Gemma Rigau and Lluïsa Gràcia for being interested in my work and for accepting to be members of the dissertation committee.

My last words of gratitude are for my family, especially my parents, Josep and Mª Dolors, who taught me all the important things about life; my brothers and sisters, Dolo, Miquel, Josep, Montse, Elisabet and Carolina, and their partners, for being there when I needed them; my nephews and nieces, Adrià, Marc, Joel, Júlia, Carolina, Arnau, and Guillem, for making me forget I was in the middle of writing this up with their fascinating little stories; my husband's family, Josep, Maria and Mari, for supporting me; my husband, David, for so many good things I cannot enumerate in a short list, and our son, Dídac, who albeit not physically visible at the beginning of the writing-up has accompanied me these last few months. *Gràcies!*

Susanna Padrosa Trias
Banyoles, May 2010

Abstract

The goal of this dissertation is to study a specific type of complex word-formation, namely compounding, and its relation to the morphology-syntax interface, with the ultimate aim of gaining a better understanding of the phenomenon. Different aspects of compounding are explored in this work, of which the main questions addressed in each chapter are outlined below.

The first chapter presents some evidence for the plausibility of a theory of grammar in which word syntax and phrasal syntax (which will be referred to as morphology and syntax respectively) are two distinct modules within a bigger syntactic module (cf. Jackendoff 1990, 1997, 2002, Ackema & Neeleman 2004), as well as evidence for the generation of compounds within word syntax/morphology. A morphological account of compounding, based on Ackema & Neeleman's (2004) morphosyntactic competition theory, is explored, tested with some English and Romance (Catalan and Spanish) compounds and contrasted with Harley's (2004, 2008b) syntactic analysis of compounds, based on Distributed Morphology (cf. Halle & Marantz 1993, Marantz 1997a, b, 2001, 2007, a.o.). The data examined in this chapter favour the morphologically-based account over the syntactically-based account of compound formation. For example, the former account can explain contrasts like *to meat-eat* and *to computer-generate*, while the latter cannot.

The second chapter starts by establishing the existence of heads in morphology and showing their crucial role in the classification of compounds. Then, the nature of the compounding elements in English and Catalan is examined, which is followed by a brief overview of some compound classifications. The most promising classification is that of Bisetto & Scalise (2005), according to which there are three overarching macro-types of compounds: subordinate, attributive, and coordinate, each being subdivided into endocentric and exocentric types. Another level of analysis is added to their original classification and the resulting scheme is applied when carrying out an exhaustive study of compounding in English and Catalan. Although initially adopted, Bisetto & Scalise's tripartite classification changes substantially in the course of the chapter. The three macro-types of compounds are reduced to one compounding type, based on a head vs. non-head relation, from which the different interpretations arise (subordinate, attributive). The existence of coordinate compounds and exocentric compounds is argued against.

The third chapter first explores Snyder's Compounding Parameter (Snyder 1995, 1996, 2001, 2002). After identifying which complex predicates must count as relevant to the parameter, its workings are considered in a few languages. The validity of the Compounding Parameter is questioned. It is concluded that a strict application of the compounding/complex-predicate parameter cannot be maintained nor can the alleged dependence of complex predicates on NN compounding. The second part of the chapter considers the possibility of a real connection between resultatives and compounding. To this end, two syntactic analyses of resultatives (Kratzer's 2005 and Mateu's 2000, 2010) are briefly reviewed. The conclusion is that compounding and resultative constructions seem to be two rather different phenomena. Finally, the question of why in some languages - like Catalan - NN compounds are productive, albeit to a lesser degree than NN compounds in a language like English, is addressed.

The fourth chapter brings together the main findings of this dissertation in a compact form.

List of Tables

List of Abbreviations

The following abbreviations are used either in the text or in the example glosses. Included are the abbreviations for some authors' names which have been frequently used in this dissertation.

A	adjective
a°	adjective (functional category)
A&N	Ackema and Neeleman
ACC	accusative
Adv	adverb
ALL	allative
AP	adjective phrase
aP	adjective phrase (functional category)
ASL	American Sign Language
ATR	attributive
B&S	Bisetto and Scalise
C	complementiser
Cat.	Catalan
CCR	central coincidence relation
CF	combining form
CL	clitic
CN	complex nominal
CP	complementizer phrase
CRD	coordinate
DAT	dative
DIM	diminutive
DM	Distributed Morphology
endo.	endocentric
exo.	exocentric
FCF	final combining form
fem/FEM	feminine
GAP	generic aspect phrase
G&F	Gràcia and Fullana
H&K	Hale and Keyser
ICF	initial combining form
IntP	intonational phrase
IPN	inalienable possession noun
IS	inflectional suffix
L	language
LCS	lexical-conceptual structure
LF	logical form
LI	lexical item
LIP	Lexical Integrity Principle
l-syntax	lexical-syntax
masc/MASC	masculine
MP	Minimalist Program
MS	morphological structure
N	noun

n°	noun (functional category)
NP	noun phrase
nP	noun phrase (functional category)
NOM	nominative
P	particle, preposition
PF	phonetic fom
PL	plural
PP	prepositional phrase
PPLE	past participle
prep	preposition
PRES	present
PRES.PPLE	present participle
PST	past
Quant	quantifier
REFL	reflexive
Rel. RHR	Relativized Right-hand Head Rule
RHR	Right-hand Head Rule
SC	Small Clause
sg/SG	singular
Sp.	Spanish
s-syntax	sentential-syntax
SUB	subordinate
TCR	terminal coincidence relation
ThV	thematic vowel
UG	Universal Grammar
UTAH	Uniformity of Theta Assignment Hypothesis
V	verb
v°	verb (functional category)
VI	vocabulary item
VP	verb phrase
vP	verb phrase (functional category)
WFR	word formation rule
WM	word marker

Chapter 1. The Morphology-Syntax Interface

In this chapter, we initially sketch Jackendoff's model of grammar (1990, 1997, 2002) (section 1.1), since the morphology-syntax interface theory to be developed in the chapter can be seen as zooming in on one of his three generative components, namely the syntactic one. Section 1.2 provides some evidence for the separation of word syntax and phrasal syntax (which will be referred to as morphology and syntax, respectively) within the syntactic component. It will also be shown that complex words cannot be formed by syntactic principles used in phrasal syntax, but must be formed by principles specific to word syntax. With this background in mind, section 1.3 presents Ackema & Neeleman's (2004) competition model between syntax and morphology, which they use to explain the existence of compounds (among other structures) in the morphological component. The competition model is applied to English and Romance (Catalan and Spanish) compounds, and the conclusion is that most of the data can be accounted for, provided a semantic constraint assumed in the model is refined. Although some evidence is provided for the generation of compounds in morphology (in section 1.2), there are also syntactic analyses of compounds available in the literature. For this reason, we felt it necessary to contrast a morphologically-based account of compounding (which we adopt) with a syntactic approach to the phenomenon. To this end, section 1.4 presents the core assumptions of Distributed Morphology, a model of grammar which endorses the view that all word formation is syntactic (cf. Halle & Marantz 1993, Marantz 1997a, b, 2001, 2007, a.o.). Framed within this model, Harley's (2004, 2008b) analysis of compounds is introduced and examined. Some implications and problems of her analysis and of the framework in general are discussed. Finally, section 1.5 contains a summary of the chapter.

1.1 Jackendoff's (1990, 1997, 2002) tripartite parallel model

Jackendoff (1990, 1997, 2002) presents a theory of grammar which clearly clashes with one basic tenet of traditional generative grammar, including its latest development known as the Minimalist Program (MP) (Chomsky 1995a and subsequent work). According to the MP, syntax is the only source of generativity, and phonology and

semantics are just interpretative components that follow from syntactic structure.[1] Jackendoff rejects such a view and proposes that syntax, phonology and semantics are three creative components, which are independent of each other, though connected by interface systems. Such an approach to grammar, known as the tripartite parallel architecture, is based on the fact that each component has its own units/primitives and principles of combination, neither of them being shared by the other components (cf. Grimshaw's 1986: 748 and Borer's 1989: 46 definition of a component).[2] Hence, the impossibility of phonological and semantic structures being read off syntactic structures. To illustrate the point, let us look at some concrete examples. Concerning syntax and phonology at the sentence level first, syntactic phrases do not have exact counterparts in phonology; that is, they do not exactly correspond to a unit in the phonological structure. Consider the following example, borrowed from Jackendoff (1997: 26, ex. 7):

(1)   a. Syntax: [a [[big] house]], [a [[[very] big] house]]

        b. Phonology: [a big] [house], [a very] [big] [house]

As can be seen, [a big] and [a very] form a phonological word but have no equivalent bracketing in the syntactic representation, which indicates that phonology cannot simply follow from the syntactic structure. In a parallel fashion, intonational phrases (IntPs), a unit of phonological structures, cannot be identified with any syntactic unit (and the example in (2) cannot be regarded as a performance error, as Chomsky 1965: 13 does):

(2)   a. Syntax: [this is [the cat [that [ate [the rat [that [ate [the cheese]]]]]]]]]

        b. Phonology: [this is the cat] [that ate the rat] [that ate the cheese]

Not any intonational phrasing is possible, though. Note that the sentence in (3a) can have two possible intonational bracketings (3b, 3c), but a break after *Children's* in (3c)

---

[1] In Chomsky's (1995b: 390) terms: "L [language] is then to be understood as a generative system that constructs pairs (π, λ) that are interpreted at the A[rticulatory]-P[erceptual] and C[onceptual]-I[ntentional] interfaces, respectively, π is a PF representation and λ an LF representation (…)". Similarly, Chomsky (2004: 107) remarks that "Φ [the phonological component] and Σ [the semantic component] apply to units constructed by NS [narrow syntax], and the three components of the derivation of <PHON, SEM> proceed cyclically in parallel. L contains operations that transfer each unit to Φ and to Σ".

[2] See Jackendoff (2007) for a comparison of Parallel Architecture with other theories like mainstream Generative Grammar (where syntax drives the derivation, cf. previous footnote) and Cognitive Grammar (where syntactic formation rules are eliminated).

is not allowed, for example (see Jackendoff 1997: 27, 2002: 118-119 for discussion of these examples and for the complete syntactic bracketing of the sentence, which has been omitted here for expository reasons):

(3)   a. Syntax: Sesame St. is [a production [of [the Children's Television Workshop]]]

b. Phonology: [Sesame St. is a production of] [the Children's Television Workshop]

c. Phonology: [Sesame St.] [is a production] [of the Children's Television Workshop]

In short, although syntactic structure does not uniquely determine phonological bracketing, there are some syntactic constraints that phonology has to obey. Similarly, there are also some phonological constraints that syntax has to observe. Whereas English is generally very strict about the adjacency requirement holding between the verb and its internal argument - in not allowing adverbs like *yesterday* to intervene for instance (4) - there are some cases in which such intervention is possible and, in fact, forced by prosodic constraints (5). Such constraints require IntPs to be of the same length and to place the longest IntP at the end preferably. In the case at hand intonation clearly constrains syntax.

(4)   a. John bought a computer yesterday.
        b. *John bought yesterday a computer.

(5)   a. ?*[John bought several expensive pieces of hardware that he's been dreaming about for months] [yesterday]
        b. [John bought yesterday] [several expensive pieces of hardware that he's been dreaming about for months]

Jackendoff (2002: 120, ex. 20)

The interaction between syntax and phonology is then mutual for Jackendoff: the different bracketing structures in each component cannot be derived from the other. In addition, the vocabulary used in each component (as can be deduced from the previous examples) is not the same. For example, phonological notions like stress, phonological

word and IntPs are unknown to syntax and, by contrast, elements like syntactic phrases and functional categories like Aspect and Tense play a role in syntax but not in phonology.

In short, syntax-phonology mismatches show that the units and principles of combination in syntax and phonology are different, with the result that neither component can be reduced to the other (for further elaboration, see e.g. Liberman & Prince 1977, Selkirk 1984, 1986, Nespor & Vogel 1986, Wälchli 2005). Given the syntax-phonology mismatches, an interface between the two systems is necessary for them to communicate: the structures resulting from the two systems must be matched up, and there are constraints regulating this interface (on this point see e.g. Jackendoff 2002: 118-119). A similar picture is obtained when the relation between syntax and semantics is considered, to which we now turn.

Although both syntactic and semantic structures have structural relations, their principles of combination are different. For example, whereas syntax has the head/complement relation, semantics makes use of the predicate/argument relation. The units used by syntax and semantics are not shared either. Syntactic categories like N and V, or syntactic phrases like NP and VP are absent in the semantic component where instead "entities like individuals, events, predicates, variables and quantifiers" are present (Jackendoff 2002: 124).

As was the case with the mapping between syntactic and phonological structures, there are also a number of syntax-semantics mismatches, two of which will be considered at the end of this section where some criticisms against Jackendoff's approach will be presented and discussed. For the time being, let us consider the examples in (6):

(6)   a. Colorless green ideas sleep furiously.
        b. Wooden turtle[3]

The sentence in (6a) (borrowed from Chomsky 1965) is well-formed syntactically but not semantically, which indicates a dissociation between syntax and semantics well-formedness. As for the phrase in (6b), it involves coercion: the default interpretation of turtle as an animal has to be changed to a more marked interpretation (to that of an

---

[3] Example (6b) is taken from Jackendoff (1997: 65).

object) to avoid semantic anomaly. Coercion is not motivated by syntax (see Jackendoff 1997: chapter 3 for other examples involving coercion). In short, although syntax may constrain semantic interpretation, the latter does not seem to be determined by syntax.

The relation of syntax to conceptual structure is then not as simple as it may seem initially. Other examples which illustrate the absence of a one-to-one correspondence between syntax and semantics are presented in (7).

(7)   a. Norbert is, *I think*, a genius.
        a'. I think Norbert is a genius
        b. An *occasional* sailor walked by.
        b'. Occasionally a sailor walked by

<div align="right">Jackendoff (1997: 38, ex. 25)</div>

In both (7a) and (7b) the italicized words are not interpreted in their surface syntactic position. As the sentences in (7a') and (7b') show, they are interpreted higher in the structure. Again, these examples indicate that the semantics and the surface syntax do not always match (see Jackendoff 1997: 33-36, 2002: 138-149 for other mismatches between syntax and semantics).

In short, the mismatches between syntactic and semantic representations seem to suggest that, like syntax and phonology, syntax and semantics are two autonomous generative systems, each with their own units and principles of combination. For the two components to be able to communicate, an interface mediates between them and constrains their relations to avoid an unrestricted interface where all imaginable relations are allowed (see Jackendoff 2002: 138f).

Not only are the types of mismatches discussed so far found at the sentence level but also at the word level (cf. Sproat 1985, Zubizarreta 1985). For example, bracketing paradoxes like *unhappier* and *ungrammaticality* are standard examples of the mismatches holding between the phonological and syntactico-semantic structures (cf. e.g. Sproat 1985, 1988, Spencer 1988, 1991). Let us consider *unhappier*. From a phonological point of view, the –*er* suffix requires that it attach to an adjective with a single foot so that *happy* and -*er* must be merged first, with the result merging with *un-* (8a). By contrast, from a semantic point of view, *happy* needs to be merged with the prefix first and with the suffix last, in order to derive the correct interpretation (not 'not more happy' (8a), but 'more not happy', i.e. 'less happy' (8b)).

(8)   a. [un+[happy+er]] [4]

       b. [[un+happy]+er]

Data from language acquisition show mismatches between the acquisition of a (morpho)syntactic form and its (morpho)phonological representation. In English, children first grasp the (morpho)syntactic properties of the agentive suffix *–er* before they learn its phonological form. That is, children give the correct interpretation of agentive nouns with *–er* like *kicker*, which suggests that the syntax and semantics of the suffix are acquired. However, when asked to produce agentive nouns, the initial forms produced by the same children are underived, with no overt suffix, until they learn that the overt realization of the suffix is *–er*. Then, because *–er* is the most productive agentive suffix in English, its use is overgeneralized to agent names that take suffixes other than *–er* (e.g. *–ist, -ian*) until children's production becomes adult-like (cf. Clark & Hecht 1982, Clark 1993, Clark 2003). The different stages in the acquisition of subject names show that the semantics and syntax of an affix may be acquired first while its overt realization (its phonological representation) is acquired later. A model of grammar like Jackendoff's can explain the dissociation between the acquisition of the (morpho)syntactic (as well as semantic) features of an affix and the acquisition of its (morpho)phonological features, whereas the same facts are hard to explain in
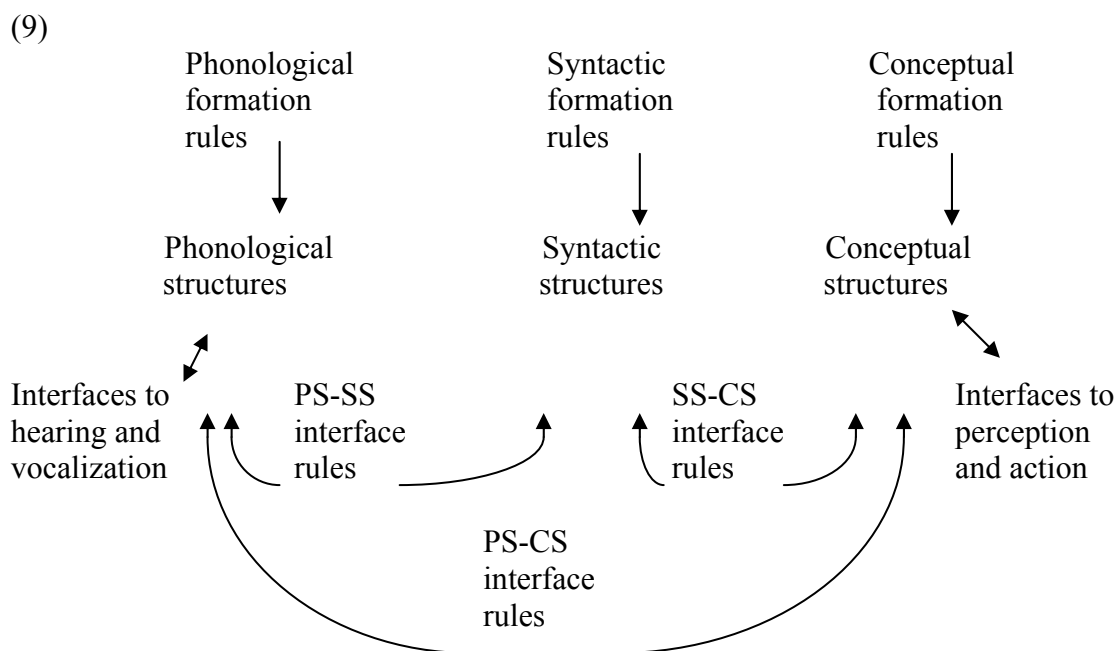
---

[4] A '+' sign has been used to signal the two elements of the affixed word. As will be seen later on, compounds in Catalan can be spelt as one word, as two words or hyphenated. The same '+' sign has been used to signal the two elements of the compound when they are spelt as one word (e.g. *cobre+llit* (cover+bed) 'bedspread'), unlike English compounds. The rest of the compounds have been written as they are conventionally spelt (e.g. with a hyphen, as two separate words), like the compounds in English. In the gloss of the Catalan compounds, though, the '+' sign has not only been used to separate the compounding elements when the compound is spelt as one word (see the example above) but also when it is hyphenated (e.g. *busca-raons* (look.for+reasons) 'troublemaker'. The '+' sign has been used for the latter case to avoid confusion: grammatical information in the gloss, written in small capitals, is separated from a lexeme by a hyphen: e.g. *pometa* (apple-DIM) 'small apple'. For the same reason the '+' sign is also used in the gloss, for example, to separate a verb and a clitic (separated by a hyphen in the spelling): e.g. *menjar-les* (eat-INF+them) 'to eat them'.

    The same glossing system has been applied to all non-English examples which required a detailed gloss to understand the phenomenon in question, with the exception of those cases in which the source was not explicit enough. In those cases, the gloss from the source has been incorporated without making any changes, with the consequence that different strategies for glossing coexist in this work.

    Note also that the terms *noun(s)*, *verb(s)*, *adjective(s)* and *adverb(s)* are usually spelt as such, but when their use is very frequent in some passages, the abbreviated forms *N(s)*, *V(s)*, *A(s)* and *Adv(s)* are used instead. The shorter forms are also used in the syntactic analysis of some (parts of) words: e.g. [[book]N sellerN]N. When doing syntactic analysis using square brackets, the grammatical information is expressed by means of a subscript, as in the previous example, instead of a hyphen. This option is chosen to avoid a cumbersome analysis. (See the list of abbreviations)

nonmodular theories. The same type of dissociation presented here can also be observed in the acquisition of synthetic compounds like *wagon puller* (cf. Clark, Hecht & Mulford 1986; see also Ackema & Neeleman (A&N) 2004: 139-144 and 154-159 for principles that constrain the mapping between morphosyntax and morphophonology and for an interpretation of the results in Clark, Hecht & Mulford 1986, respectively; cf. also A&N 2002).[5]

In short, facts like those discussed above led Jackendoff (2002: 125) to propose the model of grammar depicted in (9) (where interface systems are indicated by double arrows):

(9)

| Phonological formation rules | Syntactic formation rules | Conceptual formation rules |

| Phonological structures | Syntactic structures | Conceptual structures |

Interfaces to hearing and vocalization   PS-SS interface rules   SS-CS interface rules   Interfaces to perception and action

PS-CS interface rules

A consequence of such a model of grammar is that a lexical item (LI) is not inserted in its entirety in syntactic structure, as in the MP (Chomsky 1995a and subsequent work). That is, a LI is not inserted with all the syntactic, phonological and semantic features from the beginning, with the phonological and semantic features being inert throughout the derivation until they reach the appropriate components. Instead, Jackendoff understands a LI as the result of linking the relevant phonological, syntactic and

---

[5] See also Jackendoff (2007) for an illustration of how NN compounds in English have a simple syntax but a complex semantics (which can include multiple coercion functions, cocomposition of these functions, etc.), with the result that there are no one-to-one correspondence rules between syntax and semantics.

semantic structures in all three components (for example, by sharing a numerical index).[6] In Jackendoff's (1997: 89-90) terms:

(10)   "(…) *a lexical item is to be regarded as a correspondence rule, and the lexicon as a whole is to be regarded as part of the PS-SS [Phonological Structure-Syntactic Structure] and SS-CS [Syntactic Structure-Conceptual Structure] interface modules*. On this view, the formal role of lexical items is not that they are "inserted" into syntactic derivations, but rather that they license the correspondence of certain (near-)terminal symbols of syntactic structure with phonological and conceptual structures. There is no operation of insertion, only the satisfaction of constraints."
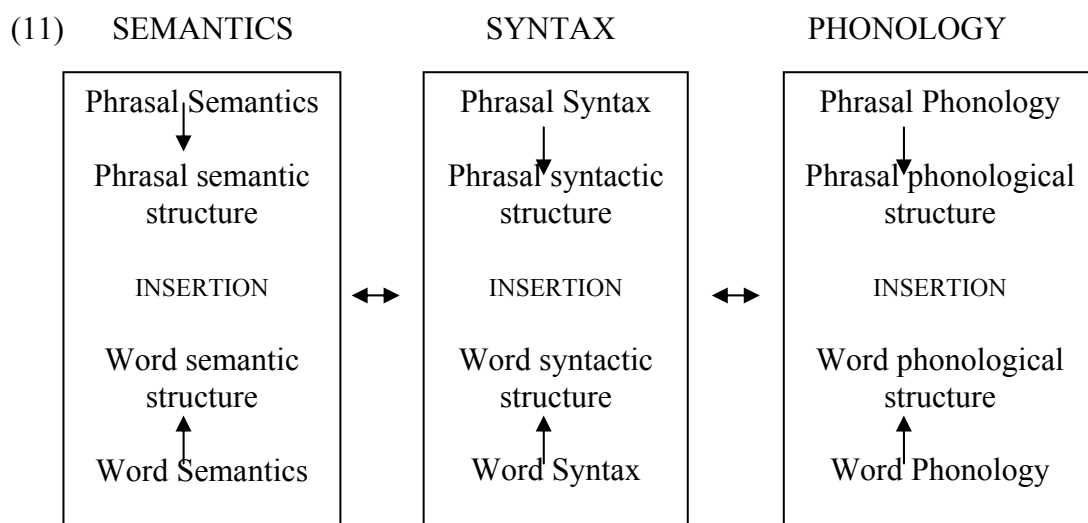
Although the phonological/syntactic/conceptual formation rules in (9) are intended to apply at both word and phrasal levels (see Jackendoff 1997: 113 for an illustrative table), it is not clear whether the units and principles of combination in each subcomponent (e.g. word semantics and phrasal semantics)[7] are the same. Jackendoff (1990) holds that they are the same (at least the "basic alphabet"), while Jackendoff's (2002) position is uncertain.[8] On this point we will follow A&N (2004), who have proposed a model of grammar similar to that in (9), with mapping principles between phonology, syntax and semantics, as shown in (11) (see also Ackema 1999a: chapter 5)[9]: they take the word and phrasal subcomponents of phonology, syntax and semantics to have their own vocabulary and principles of combination although some of them are also shared by the two subcomponents by their being inserted in larger phonological, syntactic and semantic components. The quotation in (12) makes this point clear.

---

[6] The possibility of having addresses in the form of an integer to identify lexemes in the lexicon has already been suggested in the literature (cf. e.g. Lyons 1977; compare also A&N 2004).

[7] As mentioned, 'phrasal syntax' and 'word syntax' will be referred to as 'syntax' and 'morphology' for ease of exposition.

[8] Jackendoff (1990: 18) says "Thus we can regard each component in Figure 1 [equivalent to our (9)] as divided into lexical principles (those that apply within words) and extralexical principles (those that apply to domains larger than the word level). However, the basic alphabet of primitives and principles of combination is shared by the two components." and Jackendoff (2002: 129) says "(…) phrasal syntax and morphosyntax might be regarded as semi-autonomous tiers with related but not identical organizing principles. Alternatively, they might be treated as different scales of phrasal syntax with different behaviour (…). Working out even a sketch of these alternatives is, however, beyond the scope of the present work."

[9] For other proposals similar to that of Jackendoff in the sense that there are principles mapping syntactic, phonological and semantic properties of words, see Sproat (1985) and Beard (1995), among others.

(11)   SEMANTICS   SYNTAX   PHONOLOGY

Phrasal Semantics	Phrasal Syntax	Phrasal Phonology
↓	↓	↓
Phrasal semantic structure	Phrasal syntactic structure	Phrasal phonological structure
INSERTION	INSERTION	INSERTION
Word semantic structure	Word syntactic structure	Word phonological structure
↑	↑	↓
Word Semantics	Word Syntax	Word Phonology

⟷ between SEMANTICS and SYNTAX boxes, ⟷ between SYNTAX and PHONOLOGY boxes

(12)   "(…) notions like *nominal*, *verbal*, *head*, *merge*, *c-command*, *argument*,
       complement, etc., belong to the big syntax module (…), and hence are shared by
       phrasal syntax and word syntax. In contrast, notions like *EPP*, *wh-movement*,
       and *scrambling* exclusively belong to the phrasal syntactic submodule, while
       notions like *germanic* versus *latinate* and the features that encode declension
       classes restrict merger in word syntax, but not phrasal syntax."

Like any framework, Jackendoff's is not free from criticisms. While some authors (cf.
e.g. Ackema 1999a, Gràcia et al. 2000, A&N 2004, Lieber 2004) have adopted,
extended or elaborated on the model proposed by Jackendoff, others have categorically
rejected it. In this respect, there are several works which propose a simpler syntax-
semantics interface (cf. Baker 1985, 1988, 1997, Bouchard 1995, Hale & Keyser
(H&K) 1993, 1998, 2002, Mateu & Amadas 2001, Mateu 2002). For example, Mateu
(2002) strongly criticizes two arguments, put forward by Jackendoff, for a complex
mapping between syntax and semantics. Let us consider the strength of each argument
in turn.

First, some alleged evidence that Jackendoff provides for syntax-semantics
mismatches is the fact that syntactic categories do not uniquely correspond to one
conceptual category. This is illustrated by looking at the category N (or at the NP),

which can express things (*pen*), events (*concert*) and properties (*whiteness*); and PPs, which can express places (*in the house*), paths (*to the church*), times (*in a week*), or properties (*in luck*). In the same way that a syntactic category can correspond to more than one conceptual category, the latter can also be expressed by more than one syntactic category: properties can be expressed by both NPs (*the whiteness*) and PPs (*in luck*); events can be expressed by VPs (*sing a song*) and NPs (*concert*). From this, Jackendoff concludes that there is no one-to-one correspondence between the units of syntax and the units of semantics, which suggests that the two components are independent of each other.

Adopting H&K's (1993, 1998, 2002) framework, which is in agreement with the proposal of homomorphism between syntactic and semantic structures, Mateu (2002; see also Mateu 2005 for related discussion) argues that there are three basic argument structure types, as shown in (13). Each type is associated with a particular relational semantics, thus deriving the direct syntax-semantics interface in (14).[10]

(13)   a.   x   b.   x   c.   x
          /\   /\
         x   y   z   x
                              /\
                             x   y
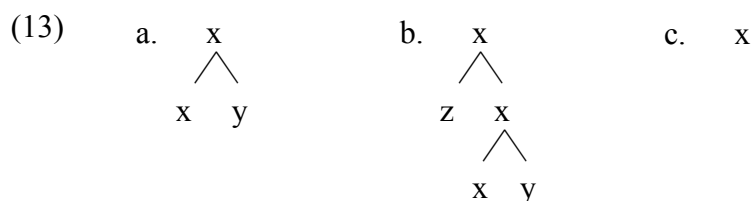
(14)   a. The lexical head *x* in (13a) is to be associated to an *eventive* relation.

       b. The lexical head *x* in (13b) is to be associated to a *non-eventive* relation.

       c. The lexical head *x* in (13c) is to be associated to a *non-relational* element.

Such homomorphism is possible because Mateu (2002: 44) understands meaning in the following way (see also Mateu & Amadas 2001):

---

[10] The data in (13) and (14) correspond to (46) and (47) in Mateu (2002: 29). Note that Mateu's three argument structure types are taken from H&K's (1998, 2002) four argument structure types. Mateu eliminates the H&K type whose morphosyntactic realization is prototypically an adjective in English, which is argued to be unnecessary. As for the rest of the types, H&K observe that in English the head (x) is prototypically a V in (13a), a P in (13b) and a N in (13c). See below for further discussion. Examples for each type are *to laugh* 'to **MAKE (x) laugh (y)**' for (13a), *to shelve books* 'to PROVIDE **books (z) with (x) a shelf (y)**' for (13b) and *cow* for (13c) (MAKE and PROVIDE should be understood as abstract verbs and the words in bold are what is structurally represented in the trees in (13)). Finally, note that (14a-b) are relational elements and together with non-relational elements (14c) constitute the primitive elements in Mateu's theory of argument structure (cf. Mateu & Amadas 2001, Mateu 2002, 2005).

(15)   "Meaning is a function of both (non-syntactically transparent) *conceptual content* and (syntactically transparent) *semantic construal*."

According to Mateu, a uniform syntax-semantics interface is possible because it is the semantic construal part of the definition, and not the conceptual content part of it, that should be taken into account in the mapping. In other words, the interface is interested in more abstract semantic notions than those that express conceptual content, the latter being full of idiosyncrasies.[11] Accordingly, a one-to-one mapping between syntactic categories and semantic notions is possible. A summary of such correspondences is given in (16). Adjectives and adverbs are not included in (16) because Mateu takes them to be derived categories, resulting from the conflation of a non-relational element with a relational one.[12]

(16)   a. Ns express non-relational elements.
        b. Vs express eventive relations.
        c. Ps express non-eventive relations.

In short, by considering a deeper level of semantics, Jackendoff's argument for many-to-many mappings between syntactic units and semantic notions has to be dismissed.

A second criticism of Jackendoff by Mateu (2002) comes from where Jackendoff (1997: 34-35) observes that the syntactic position of the internal argument can be occupied by a wide range of theta roles, such as Theme/Patient (e.g. *Emily threw the ball*), Goal (e.g. *Joe entered the room*), Beneficiary (e.g. *George helped the boys*), and Experiencer (e.g. *The story annoyed Harry*), among others. From this observation, Jackendoff concludes that the Uniformity of Theta Assignment Hypothesis (UTAH), or

---

[11] While a few other authors distinguish conceptual semantics from linguistic (grammatically-relevant) semantics, their implementation may vary in each case. For example, the linguistic meaning can be characterized syntactically, as H&K (1993, 1998, 2002) and Mateu (2002, 2005) do, or semantically, as Rappaport Hovav & Levin (1998), Lieber (2003, 2004) and Levin & Rappaport Hovav (2005) do. We will not enter into the details of each proposal since it is not relevant to the present discussion, but the interested reader is directed to the original works. Note that authors like Jackendoff (1990, 1997, 2002) and A&N (2004) do not distinguish the two types of semantics, which are both taken to be part of the semantic module of grammar.

[12] Mateu & Amadas (2001: 16) and Mateu (2002: 45-46) add that "In non-predicative contexts, Adjs typically modify non-relational elements, while Advs modify relational elements". See also Amritavalli & Jayaseelan (2003) and Kayne (2009), for the proposal that adjectives are not a primitive category but are the result of incorporating a noun into an adpositional marker.

any equivalent principle, cannot be correct. According to the proposals endorsing UTAH, some structural aspects of semantics are read off the syntactic structure: for example, identical theta-roles originate in the same syntactic position in the tree. Then, given UTAH (i.e. syntax-semantics homomorphism), the wide range of thematic roles present in the syntactic position of the internal argument in the examples provided by Jackendoff above imply that each NP coming after the V is associated with a different underlying syntactic structure (which, in turn, implies syntactic movements to derive the surface sentences). Mateu (2002: 60-61) rejects Jackendoff's argumentation by appealing to the distinction drawn in (15). That is, the UTAH should be guided by semantic construal, and not by conceptual content (for further discussion, see Mateu 1999: 3-9, Mateu & Amadas 2001: 17-21).[13]

Although we agree, in line with Mateu, that no proliferation of theta roles is desirable, we will present some evidence that speaks against a strong correlation between theta roles (as understood by H&K or Mateu) and structural positions. Padrosa-Trias (2005a, b, 2006, in press, a) adopts Reinhart's (2000, 2001) theta system and applies it to the derivation of *en*-prefixed verbs in Catalan, Spanish and English. By proposing two binary features: [*+/-c*] and [*+/-m*], Reinhart (2000, 2001) derives the Θ-roles of the 'Theta theory' found in the Principles and Parameters framework (Chomsky 1995a). Seeing that causality is crucial in thematic structures and observing that there is an overlap between the Cause and Agent roles, Reinhart labels the shared property [c], "cause change". Since agency, unlike causality, involves volition and intention, this feature is labeled [m], "mental state of the participant". By assuming two features and two possible values for each, the system generates eight feature bundles, given that not all feature bundles need to consist of two features. There is a (strong) correspondence between the clusters and the Θ-roles, of which the relevant one is given in (17):[14]

---

[13] Mateu & Amadas (2001: 19) provide the following correlations (compare Baker 1997):

   (i)   a. An Originator is the specifier of the functional projection FP.
              b. A Figure is the specifier of the non-eventive relation.
              c. A Ground is the complement of the non-eventive relation.

[14] Here are the remaining correlations established by Reinhart (2001: 3).

   (i)   [+c+m] agent
              [+c-m] instrument
              [-c+m] experiencer
              [+c] cause (unspecified for /m; consistent with agent and instrument)
              [+m] (unspecified for /c) with verbs such as *love, know, believe* (externally generated); *laugh, cry, sleep* (requiring an animate argument)

(17) [-c-m] theme/patient

Padrosa-Trias (2005b: 52) shows that "in the case of denominal Vs the [-c-m] features originate in the prefix in locative Vs [e.g. *encaixar* EN+box$_V$ 'to box' and *encaputxar* EN+hood$_V$ 'to put the hood on somebody's head'], but in the N's reinterpreted R-role in Vs of creation [e.g. *enraiar* EN+raft$_V$ 'to make/create a raft']". If this is the correct analysis, no direct mapping between thematic roles and syntactic structure is possible, contra UTAH (see Borer 2003: 40 for related discussion).

While Mateu's first argument against Jackendoff may hold, the second one does not seem to, or at least we feel more evidence is needed to support it. In addition, Mateu's (2002: 44, fn. 48) explicitly says that he will not discuss "for reasons of space" other arguments provided by Jackendoff (1997: chapter 3), which precisely present some problems for a direct syntax-semantics interface (see the examples in (6) and (7), and the original work for more problematic data). In short, despite the fact that Jackendoff should be more careful about and revise some of his arguments, there is evidence for a non-uniform mapping between syntax and semantics.

On the other hand, there are further shortcomings present in theories of argument structure which propose that the semantics can be read off from the syntactic structure. H&K's theory is a clear exponent of such a direct syntax-semantics mapping and three weaknesses of this theory, as far as we can see, will be considered to illustrate the point.

First, H&K's basic idea is that syntax is divided into l(exical) and s(entential) syntax. L-syntax, which is constrained by principles of s-syntax like head-to-head movement, is the locus where words like denominal verbs (e.g. *to shelve*) are formed. It is not clear to us why syntax is divided into lexical and sentential syntax, and then principles of s-syntax guide word formation in l-syntax. If the division is real, why is l-syntax not constrained by its own principles? Or, if word formation is guided by principles which are really syntactic (which form part of s-syntax), why is the division between l-syntax and s-syntax needed after all? To us, such a division sounds artificial

---

[-m]   (unspecified for /c) usually expressing subject matter/locative source

[-c]   (unspecified for /m) usually expressing internal roles like goal, benefactor (typically dative or PP)

and ad hoc, convenient to avoid criticisms like those in Fodor (1970). That is, H&K can circumvent Fodor's arguments against lexical decomposition (e.g. *kill* from 'cause to die') by arguing that such arguments are only applicable at s-syntax and not at l-syntax. Also, note that they do not explain how l-syntax is to be linked to s-syntax.

With Jackendoff (1997: 232), it is not clear to us either how the phonological form of *shelve* and *bathe*, among others, are brought about if the l-syntactic derivation starts out with the Ns *shelf* and *bath* respectively. Even if, for example, *shelf* and *shelve* constitute two separate lexical entries in the lexicon, with idiosyncratic phonological realization, the V is formed as a result of inserting the N at the bottom position of the l-syntactic tree and by the N moving up the tree via head movement until it gets to the V position. Given this picture, we do not understand how the change from N to V also implies a change in its phonological shape (*shelf* → *shelve*). In this respect, it is interesting to note that Carstairs-McCarthy (1992: 152, citing from Spencer 2003a: 238) reaches a similar conclusion: "the head-movement analysis, unfortunately, is incompatible with the existence of allomorphy".

Finally, we just want to point out the convenient changes that H&K's first proposal has undergone in order to be able to explain some data (exceptional data in their original proposal). Initially, Vs are derived by conflation of an N or an A into a higher empty phonological V base, thus providing it with phonological content. At this stage, H&K (1993, 1998) understand conflation as a standard head-to-head movement operation, with traces being left behind as the N moves up. Accordingly, a sentence like *We shelved the books* is expected but a sentence like *We shelved the books on the top shelf* is not, given that the position occupied by the alleged trace has been filled with new material (*on the top shelf*). When faced with such data, H&K (2002: 103) conveniently change their understanding of conflation to "it is merely the binding relation that holds between the semantic features of a V (phonologically overt now) and features of the nominal head of its complement".

To sum up, although a theoretical framework, like that of H&K, which endorses a direct mapping between semantic and syntactic structure should, in principle, be favoured for economy and transparency between interfaces, we have seen that it suffers from several shortcomings. In addition, despite rejecting some of the arguments provided by Jackendoff for non-transparent interfaces, we have seen that such non-

isomorphic interfaces are, nonetheless, necessary, a conclusion reached by considering other facts that pointed to the plausibility of a model of grammar similar to that postulated by Jackendoff or A&N. Given this brief view on plausible models of grammar, and while expecting new data and evidence to (dis)confirm our provisional conclusions, the following section is devoted to the syntactic module and its internal composition (i.e. word and phrasal submodules).

1.2 Looking inside the syntactic component

Despite the large number of works dedicated specifically to morphology and to the interaction between syntax and morphology (e.g. Anderson 1982, 1992, Aronoff 1976, 1994a, Borer 1998, Di Sciullo 1997, 2005, 2007, Felíu et al. 2006, Lieber 1992, 2004, Piera & Varela 1999, Spencer 2000, 2003a, Varela 1999, 2005, a.o.)[15], there is still no agreement in the literature as to whether morphology should be differentiated from syntax, or rather subsumed under it. This section provides some evidence for the need to distinguish them. The final outcome is that morphology and syntax constitute two distinct submodules. The fact that some vocabulary and principles apply to both subcomponents makes plausible the view that they are placed inside a bigger syntactic component, along the lines proposed by A&N (2004) (cf. 11).

1.2.1 Morphology and syntax: one component or two?

Not until the 1970s (Chomsky 1970, Halle 1973, Aronoff 1976) was morphology studied in its own right, being no longer reduced to phonology (Chomsky & Halle 1968) or to syntax (Lees 1960), as it had been in previous years. A fruitful period of work on morphology (known as lexicalist morphology, cf. Scalise 1984) followed. Some years later, though, works like Sproat (1985), Baker (1985, 1988) and Lieber (1992) questioned the idea of morphology being a component on its own, and entertained again

---

[15] Spencer (2003a: 235) notes that the denial of an autonomous morphology should not be based on the fact that there is no good characterization of the object of study, e.g. wordhood, (cf. Julien 2002), because the very same problem is present, for example, in syntax and phonology (where key notions to these fields are not fully understood either). Spencer (2003a: 236-237) further observes that "the only reasonable course of action (…) is to assume that morphology is at least partly autonomous and to investigate the principles that might be unique to it". The reasoning behind his observation is as follows: If morphology and syntax are really two different components, research on the former will uncover principles specific to morphology. If, by contrast, morphological and syntactic principles prove to be identical, nothing will be lost because the results of studying morphology and syntax separately can be put together. However, if one assumes from the beginning that morphology and syntax are the same, no one will ever know if there are principles unique to morphology.

the possibility that syntax could also explain morphological constructs. Since then, such a debate has not been settled and the dilemma still persists.[16] Recent models like Distributed Morphology (Halle & Marantz 1993, Marantz 1997a, b, 2001, 2007, Harley & Noyer 1999, Harley 2008b, a.o.) also challenge the autonomy of morphology and explain word-formation by means of syntactic principles (see section 1.4) (for other recent syntactic approaches to word formation, see e.g. Baker 2009, Borer 2008, 2009, Emonds 2006 and Julien 2002, a.o.).

If word-formation like compounding and affixation could entirely be accounted for by syntactic principles, a simplification of the grammar would result: there would be no need for a morphological component because syntax would explain both words and phrases. However desirable this picture may be, there is evidence for a morphological component, separate from the syntactic one. In fact, a number of authors have argued for the separation of morphology and syntax (A&N 2004, 2007, Bisetto & Scalise 1999, Borer 1989, Di Sciullo 2005, 2007, Di Sciullo & Williams 1987, Padrosa-Trias 2007a, b, Selkirk 1982, Williams 2007, 2008). For example, Bisetto & Scalise (1999) defend the view that morphology and syntax have their own domain, each with distinctive properties. Such a distinction permits differentiating compounds like *capo+stazione* ('station master'), which fall into the domain of morphology, and compound-like phrases like *produzione scarpe* ('shoe(s) production'), which despite sharing some properties with compounds, are nevertheless syntactic in nature (see the original work for details). Illustrative is also Di Sciullo's (2005) observation that if morphology were subsumed under syntax, additional rules would be necessary to explain morphologically specific properties.[17] In what follows some evidence will be provided for the morphology-syntax division.

There are several phenomena that indicate that morphology and syntax should be treated as two separate modules, with the consequence that words and phrases should also be treated differently. First, only words (as opposed to phrases) can delimit the

---

[16] For a good summary of how the status of morphology as a component of grammar has evolved since its origins until nowadays, see e.g. Fábregas (2006) and Val Álvaro (2006) (both in Felíu et al. 2006), and Borer (1998).

[17] In Di Sciullo's (2005: 175) terms: "One problem with this view [i.e. that of identifying morphology with syntax] is the increase of the computational load of the grammar. A single syntactic derivation for both words and phrases requires additional rules to derive word-internal properties in addition to the rules deriving phrasal properties, because syntactic and morphological properties are not coextensive".

boundaries of vowel harmony (cf. Archangeli & Pulleyblank 1994, Hualde 1998, Pilar Prieto p.c.). If words and phrases are dealt with by the same module, such diverging behaviour with respect to vowel harmony is not expected.

Second, unlike syntactic rules, morphological rules may need to refer to the phonological structure of the word before they can apply. For instance, expletive infixation requires a very specific phonological context: e.g. infixes like *bloody* and *fuckin'* can only be inserted in a word if they immediately precede a stressed syllable (cf. e.g. Siegel 1979, Aronoff 1976).

Third, a number of authors (Chomsky 1970, Bresnan & Mchombo 1995) have noticed that parts of words seem to be invisible to syntactic principles and have called such a property lexical integrity, which, if correct, signals a major difference between syntactic objects and morphological objects. Syntactic rules cannot access the internal structure of words, with the result that a morphologically complex word and a morphologically simplex word behave the same with respect to syntax. This explains why words are called syntactic atoms (cf. Di Sciullo & Williams 1987, A&N 2003): words are atoms in that their internal structure is invisible to syntax. Contrasting with words, the internal structure of phrases is visible to syntax and, consequently, syntactic rules can apply to their parts.

Whether lexical integrity is a property that follows from a principle or from the architecture of the grammar itself will not be considered here (on this point, see e.g. Fábregas et al. 2006 and Gaeta 2006).[18] For expository reasons, such a property will be called the Lexical Integrity Principle (LIP for short) and some predictions which have been claimed to follow from such a principle will be examined since, if true, they will constitute prima facie evidence for distinguishing words from phrases (i.e. morphology from syntax). Despite the controversy of some phenomena (for discussion, see A&N 2003, Anderson et al. 2006, Lieber & Scalise 2006),[19] there are facts that clearly show the validity of the LIP (for a detailed study, see Bresnan & Mchombo 1995): they show

---

[18] By attributing a particular configuration to words, Fábregas et al. (2006) derive the same effects as the property of lexical integrity. In a constructional-based morphology, Gaeta (2006) also argues for the validity of such a property by making recourse to schemas: conflated schemas (be they phrasal, affixal, etc.) do not look into each other's internal structure, thus observing lexical integrity.

[19] Notice that some threats to the LIP are only apparent. For example, by assuming that edge features are syntactic and by showing that their phonological realization is subject to the lexical properties of the word to which they are attached, Anderson et al. (2006) conclude that syntax occurs internally to words, and hence the LIP is violated. Such a conclusion is unwarranted: edge features are inflectional features (e.g. case markers), but no reason is given to treat edge features as syntactic. If they prove to be morphological, then no incursion into the LIP is necessary.

that the internal structure of words behaves differently from that of phrases. First, one piece of data which appears to be controversial will be considered; next, clear evidence for the separation of morphological units from syntactic ones will be presented.

It follows from the LIP that a word cannot contain a phrase, because it implies that the latter would have accessed the inner structure of the word, which is assumed to be impossible according to the LIP. And it is precisely to rescue a violation of such a principle that some authors have claimed that phrases are somehow fixed and lexicalised, not freely formed, when they appear in a word (see subsection 2.2.1 in chapter 2). Other authors, though, provide evidence for the opposite view: any phrase can be inserted as part of a word in Germanic languages like English, Dutch, German and Afrikaans (cf. Bauer 1983, A&N 2003, 2004, Meibauer 2007, a.o., whose position is summarized in subsection 2.2.1 in chapter 2). Once the latter view is accepted, it needs to be seen whether the LIP is necessarily violated. If we understand that a word (X^{o}) contains a phrase as such (YP), the LIP is clearly violated, but if the phrase acts as an atom inside the word, then the principle is not violated. Such a view is proposed by A&N (2004): according to their proposal, insertion is unselective and can occur between the two subcomponents in each big component (syntax, phonology and semantics). Regarding the syntactic component, A&N's proposal views as viable the insertion of morphological units (treated as syntactic atoms) into a syntactic terminal but also the insertion of syntactic units (treated as morphological atoms) into a morphological terminal.[20] In other words, the internal structure of the material that gets inserted into a different subcomponent is invisible in such a subcomponent, thus not violating the LIP (see A&N 2007: 341-349 for potential counterexamples). Once the apparent controversial data have been explained away, let us consider how morphological and syntactic units differ with respect to each other.

The LIP predicts that movement out of words is not possible, which seems to be the case as a number of studies has shown (cf. Bisetto & Scalise 1999, A&N 2003: 100-103, 2004: 35-36). In a similar vein, the LIP also predicts the impossibility of movement into words. In this respect, A&N (2003: 107-110) show that Baker's (1988)

---

[20] A phrase can appear in the non-head position of an NN compound if the features of the phrase match those of the terminal where the phrase is inserted. The phrase can be an NP but it can also be a CP or an AP, which can be seen as a problem when it comes to the matching mechanism: which features of the CP or AP are to be matched against the nominal features of the terminal where the phrase is inserted? (See Meibauer 2007: 242-243, Lieber & Scalise 2006: 22 for discussion on this point). For the insertion to work, A&N must assume that there are just very few restrictions on the non-head position of a compound (Ad Neeleman: p.c.).

arguments for incorporation, i.e. syntactic head movement of one head to another head, are not well-founded (see also Spencer 2000: 329-331 for a summary of some problems concerning Baker's 1985, 1988, 2009 approach). Their argumentation will not be reviewed here, but just notice that if syntactic movement can generate complex heads, these are expected to behave differently from morphological complex heads, a prediction borne out by the data. For example, morphological and syntactic complex heads behave differently with respect to headedness. Whereas English morphology is right-headed (Williams' 1981a Right-hand Head Rule), English syntax is left-headed (VPs, PPs) (18a vs. 18b). Regarding Catalan (and Romance in general), morphological and syntactic complex heads are not subject to the same principle of headedness either. Verb-clitic combinations, argued to be syntactic heads by Jaeggli (1986) and Borer (1984), can be right-headed (19a) or left-headed (19a') depending on the form of the verb (finite or non-finite respectively). Regarding morphological complex heads, derivation is typically right-headed (19b) and compounding left-headed (19c) (see Selkirk 1982, Scalise 1984 and also subsection 2.3.2 in chapter 2), with the result that morphology and syntax are guided by distinct factors with respect to headedness.

(18)   English
        a. [mad$_A$ness$_N$]$_N$, [computer$_N$-generate$_V$]$_V$
        b. [to the sky]$_{PP}$

(19)   Catalan
        a. [les$_{CL}$ menjarà$_V$]$_V$
           them eat-FUT.3SG
           '(s/he) will eat them'
        a'. [menjar$_V$-les$_{CL}$]$_V$
            eat-INF+them
            'to eat them'
        b. pometa
            apple-DIM
            'small apple'
        c. faldilla pantaló
           skirt   trouser
           'skort' (i.e. a type of skirt that resembles a pair of trousers)

More evidence for the separation of morphology from syntax comes from a constraint, the so-called complexity constraint, in Dutch, which syntactic complex heads are subject to, but morphological complex heads are free of (cf. Neeleman 1994, A&N 2003, 2004). A syntactic complex head which functions as a complex predicate like particle-verb or resultative-verb combinations cannot undergo further predicate formation. Consider (20).

(20)   a. dat Jan en Piet [samen werken]
            that John and Pete together work
            'that John and Pete cooperate'
       b. dat Jan en Piet zich   [kapot werken]
            that John and Pete themselves to+pieces work
            'that John and Pete work themselves to death'
       c. *dat Jan en Piet zich   [kapot [samen werken]]
            that John and Pete themselves to+pieces together work

                                   A&N (2003: 112, ex. 33; 2004: 33, ex. 29)

The syntactic complex predicates in (20a, b) do not involve recursion, which explains their grammaticality. That is, the particle verb *samen werken* and the resultative verb *kapot werken* do not undergo further predicate formation. By contrast, in (20c), the particle verb *samen werken* heads a resultative complex predicate, which is prohibited by the complexity constraint.[21]

If the complexity constraint holds for syntactic complex predicate formation, as shown in (20), Dutch verbal prefixation, suffixation and compounding must be morphological and cannot be syntactic. Unlike syntactic complex predicates, complex verbs formed by prefixation, suffixation and compounding are not subject to the complexity constraint and can head a complex predicate. The case of prefixed verbs is illustrated in (21): the verb *ver+groot* must be formed in morphology as it does not block further complex predicate formation in syntax.[22]

---

[21] An alternative approach to the complex predicate analysis for the data in (20) is the Small Clause (SC) analysis (cf. Dikken 1992, Hoekstra & Mulder 1990 but see Farrell 2005 for a recent critique of the SC analysis and McIntyre 2009 for some arguments in favour of the complex predicate analysis).
[22] See A&N (2004: 34-36, 2007: 337-339) for other differences between morphological and syntactic complex heads.

(21)   a. dat Jan de foto's [ver groot]

that John the pictures en larges

b. dat Jan de foto's [uit [ver groot]]

that John the pictures out en larges 'that John completely enlarges the picture'

A&N (2003: 113, ex. 35b, b'; 2004: 34, ex. 31b, b')

In short, the LIP has provided some evidence for generating morphological and syntactic objects in two different components. As we have seen, the LIP is, in principle, incompatible with word-formation in syntax. This explains why some proponents of syntactic word-formation (via head movement) like Baker (1988) are forced to stipulate some kind of filter to derive the same effects of the LIP (i.e. the opacity of complex X° categories).[23]

There are other factors like stranding, inheritance, referentiality, possible functions of non-heads, and derivational economy which provide more evidence for the view according to which complex words are generated by an independent morphological system (cf. Ackema 1999a, A&N 2003, 2004, 2007, McIntyre 2009, Padrosa-Trias 2007a). Such arguments supporting the morphological generation of complex words will not be discussed here for reasons of space, but the reader is referred to the works cited above.

To recap, some evidence has been provided for the separation of morphology and syntax, and for the generation of complex words in morphology and not in syntax (see also sections 1.3 and 1.4). Recall that, although separated into two subcomponents, morphology and syntax are inserted into the same component (i.e. the syntactic one), which explains some shared vocabulary and principles (cf. 11, 12). Now two more pieces of data and some positions of authors sharing the same view will be presented to reinforce the plausibility of such a model. First, notice that recursivity is a property shared by morphology and syntax, and that in both cases it is limited by extra-grammatical factors (e.g. limitations on computation and short-term memory). Traditionally, though, recursivity has been a property characterizing syntax, not

---

[23] See Ackema (1999a) for a proposal according to which morphology and syntax, although segregated, are regulated by the same principles, which, when applied to the domain of morphology, derive the effects of the LIP.

morphology, and sentence embedding has been the case par excellence to illustrate it. Regarding morphology, the limited number of prefixes and suffixes that can be used in a word, for example, has usually been taken as an indication of the non-recursive nature of morphology (as opposed to syntax). However, sentence embedding is limited by the way cognitive systems interact in the same way that morphological processes are. Consider (22).

(22)   a. This is the malt that the rat that the cat that the dog that the cow tossed worried caught ate.
        b. His great-great-great-great-great-great-great (…) -grandfather was killed in a Viking raid on Holy Island.

<div align="right">Bauer (1983: 67)</div>

In both cases, the sentences are grammatical but difficult to process. We can then conclude that recursivity is present in both syntax and morphology. [24] To convince the sceptical reader about the use of recursivity in morphology, consider (23):[25]

(23)   kindercarnavalsoptochtvoorbereidingswerkzaamhedendrukte
        child-carnival-s-parade-prepare-ing-s-work-ly-hood-PLUR-busy-ness
        'activity in connection with the preparatory work in progress for a parade at the children's carnival'

<div align="right">Battus (1985: 137, cited in Ackema 1999a: 211)</div>

Second, despite initial appearances, Ackema (1999a: 211-212) shows that the conditions to which conjunction reduction in syntax is subject, also apply to

---

[24] Other illustrative examples of recursivity in morphology and syntax are provided by Pinker (1994). First, consider examples where recursivity applies to morphology: *unmicrowaveability* (as applied to e.g. French fries), a *toothbrush-holder fastener box* and '*floccinaucinihilipilification*, defined in the Oxford English Dictionary as "the categorizing of something as worthless or trivial"', to which Pinker applies other word-formation processes: *floccinaucinihilipilificational*, *floccinaucinihilipilificationalize*, *floccinaucinihilipilificationalization*, etc. (p. 129-130). Second, consider the sentences below where recursivity has also applied ((i) with right-branching, (ii) left-branching, and (iii) multiple embedding) (p. 203-205):

   (i)   Remarkable is the rapidity of the motion of the wing of the hummingbird
   (ii)   The hummingbird's wing's motion's rapidity is remarkable
   (iii)   The rapidity that the motion that the wing that the hummingbird has has has is remarkable

[25] For some examples of how recursivity affects morphology in Romance, see e.g. Piera & Varela (1999: 4379-4380) and Varela (1999: 265) for Spanish.

morphology. One such condition is that the elided part in a conjunction must be next to the coordinator, a constraint satisfied in both syntax and morphology ((24) and (25) respectively):

(24)   a. Die muziek imponeerde haar maar __ interesseerde hem niet

          That music impressed her but   interested   him not

      b. *Zij bewondert die muziek maar hij verafschuwt __

          She admires   that music   but   he despises

(25)   a. woordintern   en   __extern

          wordinternally and externally

      b. *woordintern   en   zins__

          wordinternally and sentence

The conjunction itself in (25) is syntactic (i.e. it is a conjunction of two full words), but notice that in morphology the elision takes place within the word (see chapter 2 for discussion on the role of conjunctions inside compounds).

      Other authors have also pointed out some principles shared by morphology and syntax. For example, Ralli & Stavrou (1997) argue that morphological and syntactic expressions share the principle of binary branching. Similarly, Bok-Bennema & Kampers-Manhe (2006) believe that morphology and syntax respect the same rules and principles of Universal Grammar (UG) and consider that the morphological component is an impoverished version of the syntactic component. Di Sciullo (2005, 2007) and Williams (2007: 355) hold a similar view: the word system instantiates a subset of the relations/properties present in the phrase system.[26] (See also Bauer 2003, Varela 1999 and Piera & Varela 1999 for other properties shared by morphology and syntax). In short, several authors find some identical principles which are present in the two components, but not for this reason they want to say that there is in fact only one component. They keep the two components distinct because there are other properties which differentiate them.

---

[26] For example, Di Sciullo (2005: 16) holds that although "asymmetry is a basic property of relations in grammar and thus part of syntax and morphology", morphology is more restricted than syntax because syntax has other properties of relations (e.g. symmetry, antisymmetry) which are absent in morphology.

To conclude, it seems that a model of grammar like that of Jackendoff (1990, 1997, 2002) or A&N (2003, 2004, 2007) in which morphology and syntax are separated (by each one heading its own submodule) but at the same time tied in some way (by being inserted in the same module) is flexible enough to capture the data, some of which have been provided here. It has also been shown that complex words are formed in the morphological subcomponent. The next section is devoted to presenting the basics of A&N's (2004) competition model, which will be crucial to understand compounding in their morphologically-based account and which will be used to contrast the syntacticocentric approach to compounds of section 1.4.

1.3 Compounds in morphology

This section focuses on the core concepts of A&N's (2004) model of morphosyntactic competition in order to grasp their view on compound formation, which has to be understood as taking place in the morphological component (subsection 1.3.1). The predictions made by the competition model are tested with some English compounds first (subsection 1.3.1.1) and with some Catalan and Spanish compounds later (subsection 1.3.1.2). It will be seen that the competition model can account for most of the data examined here if one of the conditions used to establish the competition (i.e. the semantic requirement) is better characterized.

1.3.1 Ackema & Neeleman's (2004) competition model[27]

A&N endorse a view according to which syntax and morphology are two competitive generative systems, since they argue that in principle two lexical items can be combined in either component. Whether there is a syntactic or morphological preference to combine lexical items depends on the type of language. In languages like English syntactic merger will be the unmarked option, whereas in polysynthetic languages morphological merger will be the preferred option. Although A&N propose that all else being equal in languages like English and Catalan syntax wins over morphology, morphological merger is also possible under certain conditions, i.e. when there is no syntactic competitor. There is competition between syntax and morphology when both the categories merged and the semantic relation obtained are the same in the syntactic

---

[27] This subsection is drawn from Padrosa-Trias (2007b) with minor modifications. Note that the examples and tree representations are borrowed from A&N (2004).

and morphological structure. A&N (2004: 51) provide the constraint in (26), which summarizes the formal and semantic conditions just mentioned.
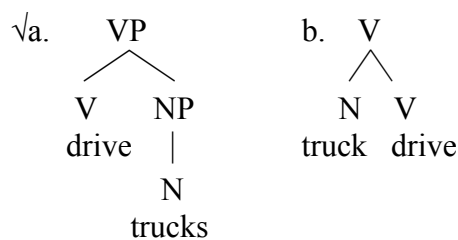
(26)   Let α_1 and α_2 be syntactic representations headed by α. α_1 blocks α_2 iff
        (i) in α_1 (a projection of) α is merged with (a projection of) β in syntax, while in α_2 (a projection of) α is merged with (a projection of) β in morphology, and
        (ii) the semantic relation between α and β is identical in α_1 and α_2.[28]

When A&N establish the morphosyntactic competition in terms of semantic identity between the morphological and syntactic structures (cf. (26ii)), they initially refer to the fact that the elements forming part of the two structures must bear the same argumental or adjunct relation in the two derivations. For example, it cannot be the case that a N is an argument of the V in the syntactic construction, and an adjunct in the morphological construction. If the latter scenario were the case, there would be no competition and the two structures would be allowed to coexist. However, it will be seen that A&N's initial proposal for semantic identity has to be refined to account for the coexistence of certain syntactic and morphological structures sharing the same argument structure. The issue of refining the semantic part of the constraint in (26) will be taken up later on (subsection 1.3.1.2), after some concrete examples have been considered. For the time being let us consider A&N's initial proposal for the constraint in (26), which is illustrated in (27) abstractly and in (28) with a concrete example.

(27)
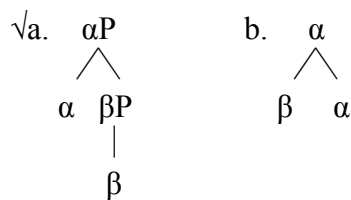
(28)   √a. They drive trucks vs. *b. They truckdrive

√a.   αP   b.   α
     /\   /\
   α βP   β   α
       |
       β

√a.   VP   b.   V
        /\   /\
       V   NP   N   V
      drive | truck drive
             N
           trucks

---

[28] The competition between morphological and syntactic structures that A&N propose could be seen as one structure blocking the other when the conditions in (26) are observed (see Embick & Marantz 2008 for different types of blocking and for their own understanding of blocking in the framework of Distributed Morphology; see also Aronoff 1994b).

In both (27) and (28) there is competition between the two structures. As for (27), the same categories, α and β, are merged and the semantic relation between them is the same in the two generative systems. Similarly, in (28) the same categories merge, i.e. a N and a V, and in both structures the N is interpreted as the object of the V. Competition is at work resulting in the syntactic structure as the winner.

As already said, morphological merger is allowed in certain circumstances, i.e. when different categories merge or the semantic relation between them is different in the two structures. A&N (2004: 52) express the difference in semantics in the following terms: "Morphological merger of α and β may result in a semantics that cannot be expressed by the result of syntactic merger of the two". (Recall that they associate having the same or different semantics with having the same or different argument structure in the two structures). To illustrate how the constraint in (26) works, let us look at some examples. First, let us consider the syntactic derivation in (29) and contrast it with its morphological counterpart in (30).

√(29)　　　NP
　　　　　／＼
　　　　N　　FP_N
　　　／＼　　／＼
　　V　N　F　NP
　drive er　of　｜
　　　　　　　　N
　　　　　　trucks

√(30)　　　　N
　　　　　　／＼
　　　　　V　　N
　　　　／＼　　er
　　　N　V
　　truck drive
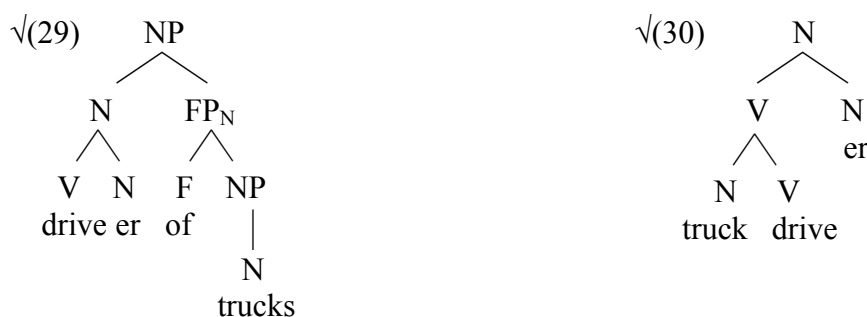
Although they both involve the same semantics (i.e. *truck* is understood as the internal theta-role of *drive* in the two cases), the merger of different categories in the two structures makes the morphological merger viable. In (29) the merger of *drive* and *–er* results in a N, which in turn merges with the N *trucks* (functional projections do not count, cf. A&N 2004: 61, but see Langacker 1999: chapter 3 for a different view).[29] In contrast, in (30), the merger of *truck* and *drive* crucially results in a V, which subsequently merges with the nominalizing suffix *-er*. To put it differently, only in (30) are *truck* and *drive* merged directly, which is what makes the morphological structure possible.

---

[29] Langacker (1999: 90) states that "*of* is a consistently meaningful element whose grammatical behaviour reflects its semantic value". In this author's view, every formal element has some meaning.

Let us consider (29) again and now contrast it with (31), another possible morphological derivation.

√(29)　　　NP
　　　　／＼
　　　N　　　FP$_N$
　　／＼　　／＼
　V　N　　F　NP
drive er　of　｜
　　　　　　　N
　　　　　trucks

√(31)　　　N
　　　　／＼
　　N　　　N
　truck　／＼
　　　　V　N
　　　drive　er

In this case, the two structures have the same categories merged. That is, in the two structures the V *drive* merges with the nominalizing suffix *–er*, resulting in a N, which is subsequently merged with the N *truck* in the two tree representations (recall from above that the functional projection *of* does not count). However, (29) and (31) differ in their semantics, which allows the existence of the morphological derivation. Whereas *truck* is interpreted as the internal argument of *drive* in (29), it is a modifier in (31). In short, (31) is only allowed iff *truck* is not the internal argument of *drive* but a modifier. The compound *truck driver* could refer to a driver of a car who has a picture of a truck on his T-shirt (cf. Lieber 2003: 250).

Focusing now on the two morphological representations (i.e. (30) and (31)), there are some arguments which favour the structure in (30) and not the one in (31) for synthetic compounding. Put differently, if *truck* is the internal argument of *drive*, the correct morphological derivation is (30) and not (31). The empirical evidence for this conclusion is based on different facts, among which there is the impossibility of inheriting internal arguments with an idiomatic interpretation: contrast the synthetic compound *ice breaker* (which must have the structure in (30)) with **breaker of the ice* (in the idiomatic reading). As a consequence, the N *truck* in (31) must necessarily have unpredictable semantics, because otherwise it would be blocked by (29), the syntactic counterpart which has the same merger of categories but has compositional semantics.

A&N adopt the general assumption that lexical storage should be as little as possible, with the consequence that only unpredictable information will be stored. Given that syntactic merger blocks morphological merger where both can apply, morphological merger must be triggered. The trigger may be related to unpredictable or

idiomatic readings of the morphological derivation. A&N specify the morphological locus of merger with the diacritic M, as in $<_M \alpha\beta>$. This will suspend the morphosyntactic competition and the morphological merger will be possible. That is the case with the root compound *colour code* in English. Contrast (32) with (33).

√(32)　　VP

　　V　　N
　　code　colours

√(33)　　V

　　N　　V
　　colour　code

The structure in (33) is possible because *colour code*, due to its unpredictable semantics, is stored in the lexicon, which gives it the possibility of being morphologically realized. The semantics involved in (33) can only be derived in syntax via the P *with*. The expression *code with colours* is not in competition with *colour code*, due to the fact that different categories merge in the two derivations. The syntactic derivation contains the lexical preposition *with*, which is absent in the morphological structure (see A&N 2004: 48-88 for the details of their morphosyntactic competition analysis).

The following subsection presents the morphosyntactic competition interacting with some English data.

1.3.1.1 English[30]

Given the competition model just outlined, if two lexical items can be combined both syntactically and morphologically, they should have different semantics, or the two derivations should involve merger of different categories. This seems to be the general picture for English. Let us consider some examples.

(34)　a. a child-molester

　　　a'. a molester of children

　　　b. a story-teller

　　　b'. a teller of stories

　　　c. the habit-forming

　　　c'. the forming of habits

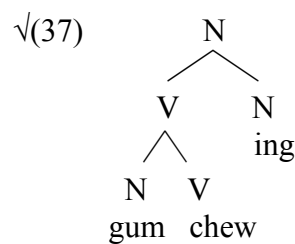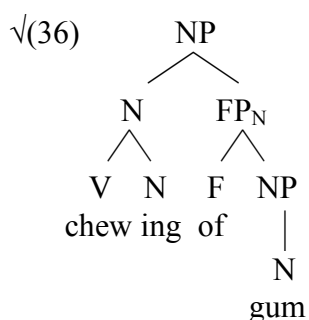[30] Parts of this subsection are drawn from Padrosa-Trias (2007b).

d. the gum-chewing

    d'. the chewing of gum

Roeper & Siegel (1978)

(35)   a. the task assignment

        a'. the assignment of the task

        b. the cake baker

        b'. the baker of cakes

        c. the trash removal

        c'. the removal of trash

        d. the housecleaning

        d'. the cleaning of the house

        e. the consumer protection

        e'. the protection of the consumer

Selkirk (1982)

The morphological merger of the lexical items in (34) and (35) is allowed in each case, because the compounds are not in competition with their corresponding syntactic counterparts. The two derivations involve merger of different categories, as can be seen in the following representations (which are the same representations as those given for *driver of trucks* and *truck driver* in (29) and (30)).

√(36)   NP   √(37)   N

        N   FP_N   V   N
                                                       ing
    V   N   F   NP   N   V
  chew ing of | gum chew
                  N
                 gum

The element responsible for having different merger of categories is a category-changing suffix (i.e. the nominalizing suffixes *–er*, *–ing*, *-al*, *-tion* and *–ment* in (34) and (35)). In the case of (36), the nominalizing suffix *–ing* merges with the underived V *chew*, the result being a N, which is crucially merged with another N subsequently; in

29

(37) the suffix merges with a compound V (i.e. *gumchew*), the result of merging the N *gum* with the V *chew*. In short, only in the morphological representation do *gum* and *chew* merge directly.

Other examples in which syntactic and morphological mergers of lexical items involve different categories are those in which the first element of the compound is not the internal argument of the base V but an adjunct which is introduced by a lexical preposition in syntax, as is shown in (38).

(38)   a. home-grown
       a'. grown at home
       b. handmade
       b'. made by hand
       c. feather-filled
       c'. filled with feathers

The preposition introduces a new category in the syntactic derivation and prevents the morphosyntactic competition, which explains why the two derivations (e.g. 38a vs. 38a') are possible.

All the examples so far illustrate that the formal condition of the constraint in (26) seems to be really at work when there are two possible structures (one morphological and one syntactic) with the same semantics. When it comes to the semantic part of the constraint, (26ii) also seems to correctly distinguish between those morphological structures which are allowed from those which are not, by comparing their semantics to that of their syntactic counterparts. Let us consider one example to illustrate how the semantic condition of the constraint in (26) explains the coexistence of (39a) and (39a').

(39)   a. to manhandle a referee
       a'. to handle a man

In this case, the two structures are allowed because the V *manhandle* is not interpreted literally, which is how the syntactic alternant in (39a') is interpreted, but roughly as 'handling roughly'. The idiosyncratic meaning attached to the compound V allows it to

be listed in the lexicon, which in turn gives it the possibility of being morphologically realized.[31]

Rice & Prideaux (1991) reach the same conclusion as A&N (2004): they observe that compound stems of the form NV rarely show up as finite verbs. They illustrate their observation with sentences like those given in (40), in which verbs used in present simple and past tense (b), infinitival form (c) and present progressive (d)[32] are ungrammatical, while those in participial constructions (e, f) and nominalizations (g) are acceptable (Rice & Prideaux, 1991: 284, ex. 3).

(40) a. They moved pianos during the music festival.
      b. *They piano-move/piano-moved during the music festival.
      c. *They used to piano-move during the music festival.
      d. ?They're piano-moving during the music festival.
      e. The piano-moving company was hired during the festival.
      f. Piano-moving is hard work.
      g. The piano-movers were well paid.

The paradigm established in (40) is what the competition model predicts to exist. On this model, the ungrammaticality of (40b, c, d) is explained because there is a syntactic counterpart with the same meaning and merger of categories. Rice & Prideaux (1991: 284, ex. 1), though, present some more controversial data for the competition model and for their own conclusion that compound verbs of the type NV hardly ever show up as finite forms.

(41) a. He lifts/lifted weights professionally.
      b. *He weightlifts/weightlifted professionally.
      c. ??He used to weightlift professionally.
      d. He's weightlifting as part of his training program.

---

[31] Even if the two derivations of (39) could be interpreted literally, one could then argue that *man* has a different function in the two structures, namely a modifier in (39a) (given that the V is still transitive and needs an internal argument present in syntax) and an argument in (39a'). Because the semantic relation between the elements merging would be different in the two derivations, there would be no competition between the two structures and both would be allowed.
[32] The speakers consulted find this sentence ungrammatical. In addition, a Google search for the progressive of *piano-move* was unsuccessful.

e. The weightlifting competition is next.

f. Weightlifting is a good complement to aerobic exercise.

g. He's a champion weightlifter.

The grammatical judgements given in (41) for (b) and (c) do not quite match those of the speakers consulted and the results of a Google search, some of which follow:

h. He weightlifts and jogs every single day to look healthy and fit.

i. He weightlifted for approximately 7 years and recently completed 4 years in the Marine Corps (…).

j. I used to weightlift and do lots of hiking in the mountains out west (…).

The finite verbs with an incorporated N (i.e. 41h, i) and the infinitival form in (41j) are predicted not to exist because they involve the same merger of categories as their syntactic equivalents. Rice & Prideaux (1991: 285-288) provide more examples which present the same problem (42a-h):

(42)   a. He bullfights for a living.

b. He lipreads because he can't afford a hearing aid.

c. Next Tuesday, they'll sightsee.

d. He bartends for a living.

e. He beachcombs every morning before work.

f. He stagemanages the company.[33]

g. He´s deerhunting regularly now.

h. As on previous Christmas Eves, they'll be carol-singing for appreciative audiences.

---

[33] This sentence as such would not constitute a real problem for the morphosyntactic competition analysis. Recall the discussion for *manhandle* in footnote 31. The word *stage* in the compound (42f) seems to be a modifier since the compound V still requires an internal object (*the company*), which contrasts with the syntactic counterpart (*to manage the stage*), in which *stage* would be the internal argument of the V *manage*. This difference in argument structure between the two derivations is sufficient to suspend competition between them. Note, though, that there is one use of the compound verb which is more difficult to accommodate within the competition model. Consider the sentence *John stagemanages for the Royal Theatre*. In this compound *stage* seems to act as the internal argument of the verb in the same way as it does in its corresponding syntactic structure.

i. I have seen a few when I was deer hunting but I have never harvested a buck in velvet.

j. Joshua, a pupil at Brighouse High School, was carol singing with his friend (…) when the attack happened in Fairfax Crescent, Southowram.

k. For the first time in many years I was carol singing last night.

To rescue the morphosyntactic competition analysis, one might appeal to the fact that some of these verbs are defective in the sense that they cannot bear past tense inflection (e.g. *bullfought*) or that new formations on the basis of some of these words are difficult to create (e.g. *mapcombs*). Having said that, one would still want to explain why some compound verbs can coexist with their syntactic counterparts if the compounded form has transparent semantics and there is no category-changing affix present in the structure. One could also try to explain the existence of such unexpected morphological constructions by referring to the fact that most of them can only express habitual action of the event. However appealing this semantic restriction may seem, it does not seem to cover all cases. Consider the verbal forms in (42i-k), which clearly make reference to a specific occasion.[34] In addition, notice that the generic reading is also available to the syntactic constructions: all compound verbs in (42) can be paraphrased as V+N with a generic interpretation (e.g. *He fights bulls for a living; They sing carols on Christmas Eve every year*).

Although there appears to be no consistent semantic difference between the morphological and syntactic structures to account for their coexistence (thus questioning the competition model), speakers do find some distinctions in meaning between the two objects. Let us reconsider the issue of habituality: although both morphological and syntactic objects can express such notions, they do so in a different way. The notion of habituality is necessarily involved in the compounds, some of which can be viewed as a sport or a profession (e.g. *weightlift, bullfight*): if you bullfight, you are a bullfighter; but if you fight bulls, you are not necessarily a bullfighter. By contrast, habituality may be expressed by the syntactically constructed phrase, but such a notion does not necessarily have to be present. Contrast the following pair of sentences: *John lifted weights this weekend, which he never does* vs. #*John weightlifted this weekend,*

---

[34] A Google search of *deer-hunt* and *carol-sing* showed that such forms can be used without an habitual reading. Three of the results are the sentences in (42i-k).

which he never does.[35] The semantics involved in the compound reminds us of Mithun's (1984, 1986) first type of noun incorporation, which refers to a name-worthy institutionalized activity. The fact that compounds can only have an habitual reading, as opposed to phrases which can also have a punctual reading, may have to do with the fact that functional categories are absent in morphological objects but present in syntactic objects.[36] The semantic differences just discussed, though, are subtle and go beyond the semantic constraint expressed in (26) (i.e. same vs. distinct argument structure). The next subsection contains some discussion of how the semantic constraint should be changed to accommodate cases which are left unexplained if the constraint in (26) is followed strictly.

Recall that the semantic differences just discussed between morphological and syntactic objects do not apply to the data in (42g-k). To account for these data formally one might also appeal to the formal condition of the constraint in (26) and try to argue that the *–ing* ending of (42g-k) is a category-changing suffix on a par with the suffix *–er* and *–ing* found in (34) and (35). There are, though, some differences between the two types of suffixes. While the latter clearly change the category of the items they attach to (i.e. they are nominalizing or adjectivalizing suffixes: e.g. *molest*$_V$ – *molester*$_N$), the role of the former is not so clear. The *–ing* suffix attaches to a V to derive another V, which together with the V *be* forms the progressive. In such cases, the *–ing* ending seems to be best treated as a functional category with no repercussions on the categorial structure of the base it attaches to.[37] If such an approach is correct, forms like (42g-k) are left unexplained on the competition model.

Other data which seem to be problematic for the competition model are given in (43): the compound and its syntactic counterpart seem to involve the same semantics and the same merger of categories (recall that the functional preposition *of* does not count), with the prediction that only the syntactic counterpart should exist, which is clearly not the case. Compare *a crew member* and *a member of the crew*.

---

[35] Thanks are due to Jon MacDonald for providing me with such data.
[36] We hope to pursue this line of research in future work.
[37] Peter Ackema (p.c.) observes that there are some derivational affixes which can be considered heads even though they are non-category changing. For example, the suffix *–hood* can be considered a (semantic) head because the semantics of *neighbourhood* is different from *neighbour*, which can only be attributed to the presence of the suffix. If the suffix *–ing* in (42g-k) could be considered a head on a par with *–hood*, then the existence of the verbal forms *deerhunting* and *carolsinging* would no longer be problematic for the competition model. Further study of such a suffix may confirm this hypothesis, but for the time being it is not clear to us how the *–ing* suffix of such forms can be argued to be a head.

(43)　animal doctor, arrowhead, bedside, bootleg, bottleneck, brain death, brain surgery, car thief, car mechanic, cookbook author, crew member, finger surgery, fingertip, horse doctor, masthead, pinhead, probation officer, roadside, sea surface, silk merchant, table leg, and tooth decay.[38]

Once each compound is compared to its syntactic counterpart, though, some patterns can be distinguished. There is one group in which the compound and the analytic form involve merger of different categories: *finger surgery* vs. *surgery **on** the/his/her finger* (*surgery of the finger*), which suspends the competition between the two structures. The same explanation applies to *brain surgery*.

There is a second group in which compounds and *of*-phrases do not have the exact same meaning (e.g. *arrowhead* vs. *head of the arrow*, *bedside* vs. *side of the bed*, *bootleg* vs. *leg of a boot*, *bottleneck* vs. *neck of the bottle*). In some cases, in addition to the compositional meaning, the compound has an idiomatic/metaphorical reading that the *of*-counterpart cannot get: *bootleg* (as in *bootleg record*) and *bottleneck* (in the sense of a narrow stretch of road, hold-up, and problem). In other cases, very subtle semantic differences exist between the two structures: for some speakers *the side of the bed* refers to an actual portion of the bed, whereas the *bedside* seems to refer to an area close to the side of the bed, but not necessarily an actual portion of the side of the bed, although it can so refer. Similar intricate semantic differences are found in other cases: *the head of an arrow* means the front part of an arrow, which happens to correspond to *the arrowhead* (made of stone, for example); however, if there is an arrowhead separated from the rest of the arrow, speakers cannot refer to it with *the head of the arrow*. In yet other cases, the two structures (e.g. *brain death* and *the death of the brain*) appear to be equivalent but cannot be used in the same contexts, thus implying a difference in semantics: in sentences like *He suffered brain death* or *Brain death followed shortly after he stopped breathing*, *the death of the brain* would appear to be interpretable as equivalent to 'the brain's death', which is not the same as *brain death*. As will be seen (more in depth in the following subsection), such delicate semantic distinctions will be necessary for the morphosyntactic competition to work.

There is a third group in which speakers prefer one form to another one. For example, speakers prefer *animal doctor, car thief, car mechanic, horse doctor, silk*

---

[38] For further discussion of these and similar examples, see chapter 2 (the subsection on nominal compounds in English).

merchant, *tooth decay* and *masthead* to their syntactic alternatives. The *of*-counterpart of the previous compounds is not neutral. As speakers pointed out to us, no one would say *I am a mechanic of cars* except in a pragmatically marked context: e.g. *What are you a mechanic of?*,[39] in which case it would mean the same as *car mechanic*. As can be seen from the previous cases, the compound tends to be the unmarked option, although in some cases opposite results are also found: some speakers prefer *an author of cookbooks*, *the surface of the sea*, *the leg of the table* to the synthetic alternatives.[40] In short, the (un)marked nature of the competing forms explains why there is no competition between the two alternatives. In the case where there is just one option available, like *probation officer*, competition cannot be established.

Finally, there is a fourth group which seems to pose a real problem to the competition model. The synthetic and analytic forms appear to be equivalent in terms of their semantics and merger of categories: consider *crew member* vs. *a member of the crew*, *fingertip* vs. *tip of the finger*, *pinhead* vs. *head of a/the pin* and *roadside* vs. *side of a/the road*. Sentences like *I hurt my fingertip* and *I hurt the tip of my finger* are treated as equivalent. Peter Ackema (p.c.) suggests that in some cases the 'of' which appears in the syntactic construction is not as meaningless as it is in e.g. *driver of trucks*, but is actually meaningful, expressing possession, or rather 'an integral part', since for example crews generally have members and fingers generally have tips, but trucks do not necessarily have drivers (see footnote 29). If this explanation can hold for the previous examples, as seems to be the case, then the small sample of counterexamples are no longer problematic for the competition model. That is, if 'of' is not meaningless, it should be taken into account when comparing which categories merge in syntax and morphology; in that case, the same categories actually do not merge (N and N in the compound versus N(P) and P(P) in the *of*-counterpart).

---

[39] For some speakers, the dispreferred option is not a matter of preference because it is simply ungrammatical: e.g. **a doctor of horses*, **a merchant of silk*. Peter Ackema (p.c.) suggests that compounds like *animal doctor* and *car thief* can be analysed on a par with synthetic compounds, i.e. as having a morphosyntactic structure [[N V] ER] (or another nominalizing AFFIX), rather than just [N N]. For example, *animal doctor* would have the morpho-syntactic structure [ANIMAL HEAL] ER] and *car thief* would have the morpho-syntactic structure [[CAR STEAL] ER]. English then has mapping rules between morphosyntax and morphophonology like 'if ER selects (a category headed by) STEAL, the phonological realization of (STEAL, ER) = /thief/ (cf. A&N 2004: 138f for discussion of such rules). This analysis would also suspend the morphosyntactic competition because different categories would merge in the compound and in the *of*-counterpart.

[40] The direction of the preference may have to do with American vs. British English, a hypothesis which needs to be confirmed.

To recap, most of the English data considered in this subsection pose no problem to the morphosyntactic competition analysis, as it was originally proposed. There is either a difference in semantics (understood in terms of argument structure) or a difference in the category merging in the two structures. However, there is a set of data (41-43) which is difficult to accommodate within the competition model if the semantic condition in (26) is not refined (see next subsection for further discussion on this point).

Next, some Catalan and Spanish compounds will be considered in relation to the morphosyntactic competition model, and conclusions similar to the ones drawn in the present subsection will be reached.

1.3.1.2 Catalan and Spanish[41]

This subsection presents two types of Catalan verbal compounds, i.e. $[NV]_V$ and $[AdvV]_V$, which together with their syntactic counterparts will also be used to further test the competition theory. The $[AdvV]_V$ type of compound is also exemplified with Spanish data.

Despite their low presence in the language (see chapter 2 and Padrosa-Trias 2007a for discussion), Catalan $[NV]_V$ compounds are already indicative of the validity of the morphosyntactic model. Given that in most cases the same categories merge, this factor will be considered only occasionally to validate the competition analysis. By contrast, the different semantics between the two structures will be the main factor taken into account to validate the competition model.

At first sight, one might argue that Catalan $[NV]_V$ compounds and their syntactic counterparts share the same semantics, a stand taken by authors such as Mascaró (1986) and Cabré & Rigau (1986), who assume that, for example, *portar a coll* (carry on neck) is the same as *coll+portar* (neck+carry), and *trencar la cama* (break the leg) is the same as *cama+trencar* (leg+break). Although at a superficial level, this generalization seems to be correct (with the consequence that A&N's morphosyntactic competition is put into question), a deeper level of analysis shows that the semantics of the two structures is not exactly identical. Let us consider some examples.

---

[41] Parts of this subsection are borrowed from Padrosa-Trias (2007a, b).

(44)   a. El   caçador   ala+trencà   els ocells.[42]
          The hunter   wing+break-PST.3SG   the birds

        b. El   caçador trencà   les ales   als   ocells.
          The hunter   break-PST.3SG   the wings to+the birds

        c. El   caçador trencà   els ocells per les ales.
          The hunter   break-PST.3SG   the birds by the wings

        d. El   caçador trencà   les ales dels   ocells.
          The hunter   break-PST.3SG   the wings of+the   birds

Because more than one paraphrase for the V *ala+trencar* is possible, one might think
that the morphosyntactic competition should predict the non-existence of the compound.
However, it will be seen that in fact there is no competition between (a) and the rest of
the structures in (44). The compound has an obligatory inalienable possession reading,
which is also present in the syntactic paraphrases of (44b, c). In the case of (44b, c), the
lexical prepositions (*als*, *per*) will prevent competition from taking place. Different
categories will merge in the morphologically and syntactically derived structures.
Regarding (44d), it cannot have an inalienable possession reading. The difference in
meaning between (44a) and (44d) will suspend the competition and hence the
morphological structure is allowed. Despite appearances, the notion of (in)alienability is
compatible with the semantic condition in (26ii), which follows directly from Vergnaud
& Zubizarreta's (1992: 596) proposal, according to which "An inalienable noun, but not
an alienable one, takes a possessor argument" (on the issue of (in)alienability, see
Alexiadou 2003 and Guerón 1985, 2003, a.o.). Translated into our examples, (44a) has
the inalienable possession noun (IPN) *ala*, which according to Vergnaud &
Zubizarreta's statement must take an argument, *els ocells*. By contrast, *les ales* in (44d)
is not understood as an IPN and hence does not take an argument but the adjunct *els
ocells*. In addition to the different role played by *els ocells*, notice that *ala* (44a) / *les
ales* (44d) also has a different role in the two sentences. The N *ala* in the NV compound
ala+trencà (44a) can only be a modifier since the compound as a whole takes an
internal argument, *els ocells* (see chapter 2 for further discussion); by contrast, the
nominal phrase *les ales* in (44d) is the internal argument of the verb *trencà*. In short, the
data in (44) can be accounted for by the constraint in (26), i.e. either by appealing to a

---

[42] Some speakers regard compounds like *ala+trencar* a little marked, which we attribute to the fact that
the habitual activity denoted by the compound is no longer a common one in our society.

different merger of categories or to a difference in semantics that has to do with argument structure (argument vs. adjunct).[43]

Let us look at more data and consider how the morphosyntactic theory fares with them.[44]

(45)   a. La Maria es trencà   la cama.
         The Mary CL break-PST.3SG the leg
         'Mary broke her leg' (Mary can be understood as an Agent or as an
           Experiencer)
       b. La Maria es cama+trencà.
         The Mary CL leg+break-PST.3SG
         'Mary broke her leg' (Mary can only be understood as the Experiencer)

(46)   a. El doctor glaçà   la sang de la Maria.
         The doctor freeze-PST.3SG the blood of the Mary
         'The doctor froze Mary's blood', 'Mary was scared stiff'
       b. Aquella notícia terrible sang+glaçà   la Maria.
         That news terrible blood+freeze-PST.3SG the Mary
         'Mary was scared stiff by that terrible piece of news'

(47)   a. En Joan porta   a coll el seu fill.
         The John carry-PRES.3SG on neck the his son
         'John carries his son on his shoulders', 'John carries his son (the manner not
           being specified)'
       b. En Joan coll+porta   el seu fill.
         The John neck+carry-PRES.3SG the his son
         'John carries his son on his shoulders'

In the case of (45), (a) can have two possible readings: one in which *Mary* is an agent, i.e. she performs the action on purpose, and another one in which she is an experiencer, i.e. Mary's leg broke by accident. Of the two possible readings, (45b) has only the

---

[43] See Brunelli (2003) for parallel examples and contrasts in Italian.
[44] Other examples can be found in Padrosa-Trias (2007a).

latter. This difference in meaning between the two structures would suspend the competition, which would explain the existence of the compound.

Similarly, in the case of (46) and (47) the syntactic derivation allows a wider range of interpretations than the morphological derivation. Again, this difference in semantics would explain why the compound is a possible derivation. Concerning (46), (a) can have a literal and a metaphorical reading, whereas (b) can only be understood metaphorically. Regarding (47), *portar a coll* can be understood literally as carrying somebody on one's shoulders and also as simply carrying somebody without specifying the manner; *coll+portar* can only have the former reading. In addition, note that the syntactic structure involves the lexical preposition *a*, which is absent in the compound, thus also explaining the coexistence of the two structures.

In short, at first sight it seems that the examples above can only be accommodated under the morphosyntactic competition if very fine-grained differences in meaning between syntactically and morphologically derived structures are taken into account (e.g. agent vs. experiencer, literal vs. figurative readings). In other words, the semantic part of the constraint given in (26) does not seem to be sufficient to suspend the competition between syntax and morphology since it only takes into consideration the argument structure of the predicate. If that is the same in the two structures, competition establishes the syntactic derivation as the winner. On closer examination, though, the examples in (45) and (46) fit into the semantic condition stated in (26). (We are putting (47) aside because the merger of different categories in the two components already suspends the competition, thus allowing the two derivations). The agent vs. experiencer readings in (45) and the literal vs. figurative readings in (46) can both be understood in terms of (in)alienability, which as explained above can in turn be understood in terms of argument vs. adjunct objects. Let us illustrate the point with the example in (46). The body part *sang* 'blood' can be understood as an IPN or as a non-IPN. In the former reading, following Vergnaud & Zubizarreta (1992), the IPN takes a semantically dependent element which is the possessor argument *la Maria*. In the non-IPN reading, *sang* 'blood' does not take an argument and hence *la Maria* is not an argument but an adjunct. While the compound (46b) can only have the reading in which *la Maria* is an argument, its analytic counterpart (46a) can have both the argument and adjunct readings for *la Maria*. At this stage we have managed to explain the different meanings in terms of argument structure (cf. 26ii), but if one of the possible readings of the syntactic structure is truly equivalent to the reading of the compound, there is still a

problem. However, there is a further difference between (46a) vs. (46b) to consider: *sang* is a modifier in the case of the compound but must be an internal argument in the case of its syntactic counterpart (see the explanation for *ala+trencà* vs. *trencà les ales* in (44) above). The same explanation explains the agent vs. experiencer readings of (45). In short, the constraint in (26) alone, without amendments, can account for the data in (45-47).

Regarding [AdvV]$_V$ compounds in Catalan and Spanish, some subtypes will be presented in what follows (for more data, see the subsection on verbal compounds in chapter 2 and Padrosa-Trias 2007b). There are some subtypes in which the categories merged in the compound are clearly different from those merging in the syntactic structure. In these cases, there is no competition between the two structures and both are predicted to exist. Consider the following example in Catalan:

(48) a. menys+tenir (less+have)

a'. tenir per menys (have for less)

'to underestimate'

In (48) the syntactic and morphological structures have different merger of categories. In the two structures a verb merges with an adverb, but the syntactic structure has an additional merger due to the lexical preposition *per*, which prevents competition between the two structures.

A relatively productive verbal compound type in Catalan is formed by the adverb *mal* 'badly' with a qualitative meaning and a verb (cf. Buenafuentes 2001-2002), which is the next type to be considered. The syntactic counterpart of *mal* in Catalan is *malament* whereas it remains the same in Spanish. Let us consider some examples.

(49) Catalan

a. mal+vendre (badly+sell) 'to sell (something) cheap'

a'. vendre malament

b. mal+gastar (badly+spend) 'to waste money'

b'. gastar malament

c. mal+tractar (badly+treat) 'to ill-treat'

c'. tractar malament

d. mal+criar (badly+bring.up) 'to spoil (sb)'

d'. criar malament

e. mal+parlar (badly+speak) 'to speak ill of somebody'

e'. parlar malament

f. mal+pensar (badly+think) 'to think badly'

f'. pensar malament

g. mal+encaminar (badly+direct) 'to misdirect'

g'. encaminar malament

h. mal+acostumar (badly+get.used.to) 'to spoil (sb)/to get sb into a bad habit'

h'. acostumar malament

i. mal+entendre (badly+understand) 'to misunderstand'

i'. entendre malament

(50)   Spanish

a. mal+vender (badly+sell) 'to sell (something) cheap'

a'. vender mal

b. mal+educar (badly+raise) 'to spoil (sb)'

b'. educar mal

c. mal+gastar (badly+spend) 'to waste money'

c'. gastar mal

d. mal+tratar (badly+treat) 'to ill-treat'

d'. tratar mal

e. mal+interpretar (badly+interpret) 'to misinterpret'

e'. interpretar mal

The compounds in (49) and (50) are a real problem for the morphosyntactic competition if the semantic condition in (26) is not refined. In other words, a verb merges with an adverb which can be taken as a modifier both in the compounds and in their syntactic equivalents, the result being that the same lexical items and argument structure are shared by the two components. If the semantic condition is not refined beyond identity of argument structure, all the examples in (49-50) are problematic for the competition model. A&N are aware that by the constraint alone, the existence of some

morphological objects cannot be explained. They are then forced to assume that, despite having the same merger of categories and the same argument structure, some syntactic and morphological constructs can coexist because the two structures diverge in their semantics in some way: for example, the syntactic structure is interpreted literally while the morphological one figuratively[45] or because, despite having the exact same meaning, the morphological structure is used for official documents, i.e. for a more formal register, while its syntactic counterpart is used for more informal situations, namely the two constructions are used in different registers.

Let us reconsider the examples in (49) and (50) in the light of the readjustment of the semantic constraint (as initially proposed). As will be seen shortly, the data in (49-50) can be divided into two subgroups.

One could argue that some syntactic derivations allow a wider range of interpretations than the morphological derivations. In other words, the semantics of the morphological construct can be viewed as a subset of the possible set of interpretations associated with the syntactic derivation.[46] That is the case of *mal+vendre* (49a) and *mal+vender* (50a), the former illustrated in (51). Compare the semantics of the following sentences.

(51) a. Els propietaris van mal+vendre el cotxe.

 'The owners sold their car cheap'

---

[45] A&N (2004: 84) exemplify the coexistence of syntactic and morphological derivations by giving some verb-particle constructions in Swedish (from Holmes & Hinchliffe 1994: 321). The syntactic derivation is interpreted literally while the morphological structure is interpreted figuratively.

 (i) Jag bryter av kvisten.
   I   break off the+branch
   'I break off the branch'
 (ii) Jag avbryter   samtalet.
   I   off+break the+conversation
   'I interrupt the conversation'

[46] This subset relationship could at first sight be related to Kiparsky's (1997: 482-483) distinction between those verbs that are named after a thing, which involve a canonical use of the thing, and those that are not named after a thing and can have interpretations other than the one just mentioned. Contrast the semantics between *to saddle a horse* and *to put the saddle on a horse*. The denominal verb can only mean that you have put the saddle on in such a way that now you can ride it. Although this interpretation can also be derived from *putting the saddle on a horse*, this expression can also have other interpretations (e.g. the saddle is on the horse but you cannot ride the horse because the saddle is not fitted in the appropriate/canonical way). The denominal verb (the morphological derivation in our case) seems to have prototypical/canonical semantic features associated with it, not present in the analytic variant (the syntactic derivation in our terms). However, as Harley (2008a) notes, Kiparsky's 'Canonical Use Constraint' is also applicable to syntactic constructions. Contrast, for example, *John is going to school* (for educational purposes, with a generic use of the noun *school* and hence with no particular referent being picked out) with *John is going to the school* (with *school* referring to a particular building).

b. Els propietaris van vendre malament el cotxe.

'The owners sold the car {cheap / in a bad condition / in an unprofessional manner (e.g. maybe the seller was swearing)}'

Even then, one would like to know why the two derivations are not competing for the shared reading. Peter Ackema (p.c.) suggests that the syntactic structure may be semantically underspecified (something like 'sell in a way that is not good in some sense or other'), the 'cheap' reading just being compatible with this underspecification, while the compound is semantically specified as meaning 'sell cheaply' only (see below for another suggestion).

As for the rest of the cases in (49) and (50), although the two expressions seem to have identical semantics, they cannot be freely exchanged in some contexts, which means that there is a semantic difference between them, not visible at first sight. Consider the following Catalan sentences.

(52)   a. Quan surts sempre {mal+gastes (badly+spend) / gastes malament} els diners.
         'When you go out you always waste your money'
        a'. Ell va {mal+gastar (badly+spend) / #gastar malament} la joventut.
         'He wasted his youth'
        b. No {mal+tractis (badly+treat) / tractis malament} el nen.
         'Don't ill-treat the child'
        b'. No {#mal+tractis (badly+treat) / tractis malament} la taula que és molt cara.
         'Don't damage the table because it is an expensive one'

Again, one would like to know why the reading shared by morphology and syntax is allowed under the competition model. Note that one might argue that the data in (49) and (50) are not a real problem for the competition model on the following grounds. In principle, one would expect the same range of interpretations in the two derivations if both have compositional semantics. If the compound has only one particular reading out of the possible readings the syntactic derivation has, this can be taken as evidence for the listing of the compound, which will be due to its idiosyncratic nature. Bear in mind that accepting such argumentation implies that a really fine-grained semantic analysis is needed for the morphosyntactic competition theory to work, which clearly shows that

identity vs. distinctness of argument structure (A&N's 2004 initial proposal for the semantic constraint in (26ii)) is not sufficient.

Finally, there are some compounds which at first glance seem to be indistinguishable from their syntactic counterparts as far as their semantics is concerned. On closer examination, though, there are semantic differences between the morphological and syntactic constructions similar to those found with the examples in (52). For example, there are some contexts in which the compound is not allowed but its syntactic counterpart is. This is exemplified by the Catalan compound *mal+entendre* 'to misunderstand' in (53a) and the Spanish compound *mal+interpretar* 'to misinterpret' in (53b).

(53)   a. Un problema matemàtic es pot {*mal+entendre (badly+understand) / entendre malament}.

      'A maths problem can be understood wrongly'

    b. La orquesta {*mal+interpretó (badly+interpreted) / interpretó mal} la sinfonía

      'The orchestra interpreted the symphony wrongly'

To sum up, the data in (49-53) show that the argument structure between the morphologically and syntactically derived expressions is the same, and still the two objects are allowed to exist (contra A&N's original proposal, according to which only the syntactic object should exist). Accordingly, a change in the original definition of A&N's (2004) semantic constraint is then needed, as they themselves acknowledge by pointing out that, for example, the contrast between formal and informal registers must be enough to suspend the morphosyntactic competition.

In short, the competition analysis between syntax and morphology, as put forward by A&N (2004), seeks to explain the coexistence of syntactic and morphological structures, which is explained either by appealing to a difference in the semantics of the two structures understood in terms of argument structure or by a different merger of categories in the two constructions. It has been shown that the morphosyntactic competition theory can explain most of the data examined here provided the semantic condition of the constraint in (26) is refined. A&N show that, in addition to diverging argument structures, a syntactic and a morphological structure can coexist if the two structures diverge in their semantics in some way. We have found out

that subtle semantic distinctions between the two structures must be taken into account for the competition model to work. For example, the notion of habituality (41, 42), literal vs. figurative readings (43), marked vs. unmarked interpretations (43), possession/integral part readings (43) and different contextual uses (49-50) should count as relevant enough to allow listing of the morphological structures and allow their existence alongside syntactic variants. That is, features like habituality should be taken as idiosyncratic, i.e. as having unpredictable semantics that must be listed, which gives the possibility of specifying morphological realization. One should know exactly, though, the extent of the difference in semantics (i.e. the degree of idiosyncrasy) between the two structures. Otherwise, the theory cannot properly predict which morphological structures are allowed in the language. More data should be taken into account to further assess the morphosyntactic competition theory. To this end, the Catalan and Spanish [NA]$_A$ compound could be contrasted with its syntactic counterpart (e.g. Catalan *un noi cama-llarg* (a boy leg+long) 'a long-legged boy' with *un noi llarg de cames* (a boy long of legs), which for space reasons is not considered in this thesis (cf. García Lozano 1978, Cabré & Rigau 1986, Mascaró 1986, Gavarró 1990b, Rainer & Varela 1992, Gràcia & Fullana 1999, 2000, Gil Laforga 2006, Padrosa-Trias 2008).

On the competition model presented so far, a compound is then allowed to exist if it has no competitor in syntax: the existence of compounds is accounted for by having a different semantics or a different merger of categories from the syntactic counterpart. Such a view will be contrasted with a syntactic account of compounding, which will be presented in the next section.

1.4 Compounds in syntax[47]

Although some evidence has already been provided for the generation of complex words, compounds included, in a morphological component separate from the syntactic one (cf. section 1.2), let us consider how a syntactic approach to compounding fares with the data described earlier. More specifically, this section is devoted to presenting the core assumptions of Distributed Morphology (DM), a model of grammar which endorses the view that all word formation is syntactic (subsection 1.4.1). Within this syntactically-based framework, Harley's (2004, 2008b) analysis of compounds will be

---

[47] This section builds on Padrosa-Trias (2009a, b).

introduced and examined (subsection 1.4.2). Finally, some discussion about five claims made in DM work will follow (subsection 1.4.3).

1.4.1 Distributed Morphology (DM): the essentials

The main goal of this subsection is to present the primary theoretical assumptions of DM[48] and, to a minor extent, some implications and problems of the framework. (This second goal will be taken up and further extended in subsection 1.4.3). The overview of the model is mainly based on Halle & Marantz (1993), Marantz (1997a, b, 2001, 2007), Harley & Noyer (1999), and Embick & Noyer (2007). Such an exposition of DM will make it evident that a morphological account of compounding like that proposed by A&N is superior.

According to the framework of DM, there is a unique generative component, namely syntax, which is responsible for both word and phrase structure. Consequently, there is no component specifically designed for word formation, neither a morphological component (but see footnote 59) nor a generative lexicon, for example. In fact, DM denies the existence of a lexicon and the properties traditionally associated with it are here distributed in various components, which gives rise to the name of 'Distributed Morphology' (for anti-lexicalist arguments, see e.g. Marantz 1997a, b).

The syntax manipulates terminals which can contain two types of morphemes: abstract morphemes and roots (symbolised by $\sqrt{}$).[49] The former are bundles of universal grammatical (morphosyntactic) features (e.g. [Past]), and are related to functional, closed-class categories, while the latter are complexes of language-specific phonological features (for further discussion, see below), are assumed to be category neutral (e.g. $\sqrt{CAT}$), and are related to lexical, open-class categories.[50] Harley (2008b: 4) understands roots "as instructions to access certain kinds of semantic information, which may vary depending on the morphosyntactic context of the Root in question". Roots need to be categorized by a functional node containing categorial information (i.e. n°, a°, v°), a requirement which is defined by Embick and Noyer (2007: 296) as follows:

---

[48] Although properties like Late Insertion and Underspecification of Vocabulary Items (to be presented below; cf. e.g. Harley & Noyer 1999) are usually taken as distinctive properties of the framework, this is a controversial claim. Williams (2007: 359) holds that the only property that is distinctive about DM is the fact that "Phrases are built (directly) out of morphemes, with no intervening notion of *word*", which he sees as problematic, and that properties like Late Insertion, Underspecification and competition (among others) have been present in earlier analyses and hence are not unique to DM.
[49] Notation taken from Pesetsky (1995).
[50] Harley & Noyer (1998, 2000) talk about f-morphemes and l-morphemes.

(54)   Categorization Assumption

"Roots cannot appear without being categorized; Roots are categorized by combining with category-defining functional heads".

The same root can adopt different realizations, which are subject to the syntactic context in which it occurs. The root √DESTROY appears as *destruct* in the context of an *n* head and as *destroy* in the context of a *v* head (cf. Marantz 1997a; but see Marantz 2001 for a different proposal in which the root is further decomposed, a proposal summarized in subsection 1.4.3).

A tree structure, which is derived by syntactic operations like Merge and Move, is sent to LF and PF (Chomsky 1995a and subsequent work). On the way to PF, terminal nodes can undergo some readjustment operations (e.g. fission), before they are given phonological content by insertion of Vocabulary Items (VIs).[51] Regarding VIs, they can be underspecified in relation to the syntactic context in which they can be inserted (see Embick & Noyer 2007: 299-300 for some examples of syncretism, the result of VIs being underspecified; see Bonet 2007 for some problems related to syncretism) and insertion of VIs occurs in a competitive fashion. There is no agreement as to whether competition affects only abstract morphemes (e.g. Harley & Noyer 2000, Embick & Noyer 2007) or both abstract morphemes and roots (e.g. Harley 2008b). Competition is resolved by appealing to the Subset Principle (Halle 1997) or to the Elsewhere Principle (Kiparsky 1982), which explains the choice of one VI over another one when there is more than one candidate for insertion. Halle (1997) defines the Subset Principle in the following way (definition taken from Embick & Noyer 2007: 298):

(55)   Subset Principle

"The phonological exponent of a vocabulary item is inserted into a position if the item matches all or a subset of the features specified in that position. 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)".

---

[51] A VI is the relation between a phonological expression and the context in which it can appear. VIs provide the terminal nodes with phonological content.

DM endorses Late Insertion[52] as one of its core assumptions, but there is no agreement as to what its domain of applicability is (i.e. whether it should include roots or not, apart from abstract morphemes). Three positions can be distinguished. First, there is the position of Marantz (1997a), who is undecided as to whether roots come with phonological expression from the computational system (from the beginning of the derivation) or their phonological realization is inserted post-syntactically, at Spell-Out. Marantz's view can be summarized as follows (p. 204):

(56)   "It is (…) an open question how much information about roots is present in the narrow Lexicon (…)".

A second position is that Late Insertion applies to both abstract morphemes and roots (Halle & Marantz 1993, Marantz 1997b, Harley 2008b, Harley & Noyer 1999), as can be seen in Harley & Noyer (1999: 3):

(57)   "(…) the phonological expression of syntactic terminals is *in all cases* provided in the mapping to Phonological Form (PF)". (italics: ours)

Marantz (1997b: 20-22) argues for this second view on the basis of two arguments, one of which is that there is no principled distinction of how the two objects (i.e. abstract morphemes and roots) interact with the computational system. The second argument is as follows: if operations like contextual allomorphy and impoverishment are taken as a diagnostic of Late Insertion, then some roots are necessarily inserted late (see the discussion of *raise/rise* below). As far as we can see, though, there is a conflict between the DM architecture and Late Insertion of roots. If VI insertion occurs at PF, as DM claims, then it is not clear to us how the choice of VIs like *cat* and *dog* is made if the only information available from the computational system at this point is [+count singular noun], and Encyclopaedic information - which is assumed to help in the choice - is accessed only after the derivation has reached PF and LF (see Harley & Noyer 2000:

---

[52] The operation of Late Insertion implies accepting some version of the Separation Hypothesis (Beard 1995). Notice that DM and the model proposed by A&N (2004, 2007) show no difference in this respect. Both models endorse some kind of Separationism: syntactic structure is separated from phonological expression.

351-352 for related discussion).[53] Marantz (1997b: 21-22) makes it explicit how a derivation proceeds in DM:

(58) "Late insertion involves making a specific claim about the connection between LF and semantic interpretation. LF can't by itself be the input to semantic interpretation. If "cat" is inserted in the phonology at a node at which "dog" could just as well have been inserted (…) then the phonological representation, specifically the choice of Vocabulary items, must also be input to semantic interpretation. (…) Where "dog" and "cat" share all features relevant to Vocabulary insertion at LF, they tie for availability for insertion at a node and *the decision to pick one over the other is open for semantic interpretation, using the Encyclopedia*. Encyclopedic knowledge is knowledge about complete derivations and representations, with phonology and syntax included". (italics: ours)

The claim of Late Insertion for roots is questioned by Marantz (1997a) himself when he presents pairs like *raise/rise*, which according to him belong to the same root √RISE but which are realized differently depending on the type of verbalizing head (an agent projecting aspect-1 head: *raise* vs. a non-agent projecting aspect-2 head: *rise*). If, as DM holds, readjustment operations on terminal nodes take place before insertion of VIs, then some questions arise. If the root (in Syntax) is not provided with any kind of phonological features, how can one know that the result of merging aspect-1 head with the root will give *raise* and not *kill* (vs. *die*), for example? Even if the root is provided with an index that links the root to a particular phonological realization, the link to the VI will not be available until after the readjustment rules have operated and these rules are blind to what the index refers to. (Another problem associated with the *raise/rise* alternation will be discussed below, after the phase-based theory approach to word-formation has been presented.)

To resolve the problem just discussed, one could appeal to Harley & Noyer's (2000) distinction between grammatical well-formedness (if the licensing conditions of VIs have been satisfied by their being inserted in the appropriate syntactic structure), which is related to structural meaning, and pragmatic anomaly, i.e. interpretative

---

[53] Although couched in a different theoretical framework, Scalise et al. (2005: 146 and 147, fn. 8) argue for the need of accessing encyclopaedic/pragmatic information in word-formation processes.

anomalies (whether or not speakers' encyclopaedic knowledge permits the felicitous use of VIs). In short, sentences can be ungrammatical for structural reasons, and grammatical but deviant due to speakers' encyclopaedic/real-world knowledge. Let us now see some of the repercussions of Harley & Noyer's division.

In agreement with DM, Harley & Noyer assume that vocabulary insertion takes place at PF. They also assume that a VI is licensed in a particular syntactic structure: each VI has licensing requirements (e.g. a VI may be listed with a [+cause] feature, which means that it needs to be inserted as the complement of a [+cause] head), which if compatible with the syntactic structure in which it is inserted, will result in grammatical well-formedness. This, nonetheless, does not imply that the result is free from semantic anomaly, which depends on one's real-world/encyclopaedic knowledge. Let us illustrate the division between grammatical and encyclopaedic knowledge with the nouns *cat* and *dog* mentioned above. Our grammatical knowledge allows either noun to be inserted in a sentence structure, but then, depending on the choice, the resulting sentence may be pragmatically anomalous, i.e. due to our encyclopaedic knowledge. In other words, the choice between *cat* and *dog* may not have syntactic repercussions but pragmatic ones. The same explanation can be applied to *raise/kill* and *rise/die*.

However, if Harley & Noyer's division is followed, many more cases of semantic anomaly are expected: since encyclopaedic information is not accessed until VI insertion has taken place, at PF there may be many VIs compatible with a given syntactic structure (e.g. *cat/dog/…*, *raise/kill/…*, *rise/die/…*, among many others). Indeed, speakers sometimes use one word when they mean another one, which on many occasions is due to the fact that the two VIs belong to closely related semantic fields. However, in DM the wrong use of one VI for another one cannot be explained by appealing to the fact that the two VIs are semantically related since encyclopaedic information comes after the choice of the VI has been made, a difficulty in the DM architecture, as far as we can see.

Even if Chomsky's (2000, 2001, 2004, 2005, 2007) phase theory is incorporated into the structure of grammar assumed in DM, as Marantz (2001, 2007) explicitly does, the problem just discussed concerning Late Insertion is still present. Chomsky holds that syntactic computation is cyclic, constrained by locality domains, the so-called phases. Every cycle is determined by merger with a phase head, which in turn triggers the

transfer of a portion of syntactic structure to the interfaces, i.e. which is interpreted semantically and spelled-out phonologically. Chomsky (2001 and subsequent work) relates phasehood to two phase heads: C and v* (the loci of φ–features). Phasehood is equated to linguistic objects with full semantic interpretation: vPs are associated with events and CPs with propositions.[54] Marantz translates this understanding of phases at the sentence level into the word level and argues that at this level phases are triggered by categorial information, namely the "little x" heads, in his terms. On this view a phase is constituted when a root is categorized (the first phase in the word) and when an already-categorized root changes category (subsequent phases in the word).[55,56] Once a phase is created, it is sent to the interpretive components.

(59)
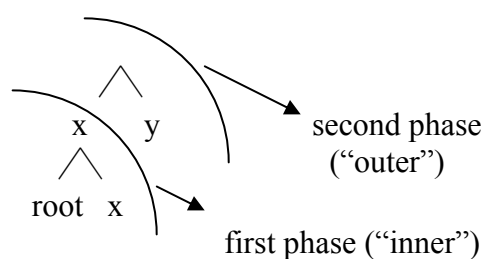
For Marantz there is a crucial difference between the first phase above the root and higher phases: the first phase is the domain where the merger of a functional head with a root may result in a special (unpredictable) phonological and semantic outcome, whereas subsequent phases are characterized by having predictable semantics and phonological form. The properties associated with each type of phase (first vs.

---

[54] A thorough analysis of Chomsky's phase theory is not intended here. See Gallego (2007, 2008) for a good summary of Chomsky's understanding of phases and for some discussion of phases as applied to Romance languages. The reader is also referred to Chomsky's original work (2000, 2001, 2004) to see, for example, how his notion of Agree in the Probe-Goal configuration has changed from deleting uninterpretable φ–features (e.g. [gender], [number]) to giving them a value, since they are now assumed to be unvalued when they enter the derivation.

[55] For a non-DM-based approach to phases as applied to words, see Di Sciullo (2005, 2007). Roughly, Di Sciullo's morphological hierarchy includes three semantic layers of projection, i.e. the predicate-argument layer (the A-Shell), the modifier layer (the Asp-Shell, which is in turn divided into internal and external to capture the Internal/External Prefix Hypothesis of Di Sciullo 1997: I-Asp and E-Asp), and the operator layer (the Op-Shell). Given this layered structure, Di Sciullo holds that phases pertain to the Asp-domain and the Op-domain. Within the Asp-domain, the phase is E-Asp, with the result that its complement (I-Asp) is sent to the interfaces and is no longer accessible for the derivation of the word. This explains, for example, that internal prefixes (directional, locational: Fr. *em+porter* 'to take away') cannot be iterated whereas the external ones (iterative, inverse: Fr. *re+re+faire* 'to redo') can. The reader is referred to the original works for further details.

[56] Chomsky and Marantz conceive the internal structure of phases differently. For Chomsky (2000, 2001) a phase must include a Probe (the phase head), a non-phase head and a Goal, whereas for Marantz (2001, 2007) there is only a phase head and a non-phase head.

subsequent) capture the properties typically associated with lexical vs. syntactic word formation (see e.g. Wasow 1977) or inner vs. outer word formation (Dubinsky & Simango 1996), which Marantz (2007: 5) summarizes as follows:

(60) Inner vs. Outer Morphology (Dubinsky & Simango (1996) et al.)

	Inner Affixation	Outer Affixation
Regularity	Potential special form and special meaning	Predictable form and predictable meaning
Selection	Attaches inside morphology determining lexical category	May attach outside morphology determining lexical category

Given this view of phases within words, the issue of Late Insertion remains unresolved. When a functional category (a little x) is hit, a phase is created and is sent to the interfaces, PF and LF. At PF, a choice of a VI needs to be made since VI insertion takes place at this level, but note that in order to choose the correct VI, Encyclopaedia information must be accessed but such access does not occur until after PF and LF, still an incongruent picture. An index may solve the problem (i.e. when a morpheme with an index is sent to LF and PF, it is given a specific interpretation and pronunciation), but not in all cases (see the discussion of the *raise/rise* alternation above). The uncertainty about (non-)late insertion may disappear, if computation and access to Encyclopaedia take place simultaneously. [57]

Although digressing from the question of Late Insertion, it is revealing to show some consequences of the phase-theory as applied to the domain of words. First, note that Marantz's understanding of phases at the word level predicts that a non-first phase head is not expected to give rise to special phonological forms. However, the suffix *–ment* '-ly' in Catalan seems to speak against such a statement: it is a regular suffix semantically, attaches to already categorized roots (to adjectives) but has some special phonological effect. Suffixation with *–ment* involves a stress-shifting process: the

---

[57] Alternatively, if Encyclopaedia were part of the Lexicon, then the puzzle could be solved. Note, though, that this position cannot be taken in DM since it denies one of their basic claims, namely that there is no Lexicon in the way it was understood in lexicalist theories, for example (cf. e.g. Lieber 1981).

primary stress of the adjectival base to which –*ment* attaches moves to the suffix and the base is given secondary stress. In addition, an inflectional element (a feminine morpheme according to some authors) is inserted between the base and the suffix: *absurd-a-ment* (absurd+a+ly) 'absurdly', *alegr-a-ment* (happy+a+ly) 'happily' (cf. Mascaró 1986: 71; see chapter 2 for further discussion on the suffix –*ment*; Teresa Cabré: p.c. holds that it is generally assumed that –*ment* behaves like a word and that it forms part of a compound). Such behaviour is not expected from a non-first-phase head. (Nor is the behaviour of some already categorized roots in Hebrew which give rise to non-compositional meaning in compounds (i.e. involving non-first-phase heads), cf. Borer 2009).

The *raise/rise* alternation also presents a conflict with the distinction of first phases vs. higher phases in the word. Recall from above that the transitive *raise* comes from an agent projecting aspect-1 head merging with the root √RISE. Following Marantz, we conclude that the functional head projecting the agent must constitute a first phase since it verbalizes the root, is associated with a special phonological form, and can have a special meaning ('to raise animals'). However, this conclusion is not in agreement with Marantz's (1997a: 208) claim that the agent-projecting head is a barrier for special meanings. That is, the presence of an agent in the structure prohibits it from having special semantics, an unsolvable contradiction in the case of *raise*, as far as we can see. Williams (2007: 360-361) presents other similar problematic examples: e.g. idioms like *The **cat** has got your tongue* and *The devil made **me** do it* contain an agent (*cat* and *me*) and presumably the *v* is present in both cases, an irreconcilable result according to Marantz's predictions (see Borer 2003: 59-63 for other problematic data).

Finally, another issue which also merits further study is the fact that in some languages a phase in Marantz's sense cannot be pronounced unless some inflectional morphology is added (e.g. tense and agreement in the case of verbs). Marantz (2007: 6) hypothesizes that for those languages, the phase is still sent to PF and LF but that it is "incapable of being uttered by itself". To explain such a phenomenon, "language-particular features of functional heads and morphological vocabulary items" (p. 7) are appealed to, not a completely satisfactory answer.

After examining briefly how Late Insertion (as applied to roots) would fare in a phase-based approach to word-formation, we conclude that no easy solution seems to be

available, which leads us to the third stance on Late Insertion, which holds that insertion does not apply to roots, only to abstract morphemes:

(61)   "(…) functional heads do not have phonetic content in the syntactic derivation. (…) By contrast, we assume Roots to be present with all their features throughout the derivation, with no such insertion process."

        "Because Roots are not subject to late insertion (…)".

                                                    Embick & Noyer (2007: 296 and 296, fn. 9)

This is the view we find most plausible. To our mind, roots must come with some phonological features specified from the beginning so that at Spell-Out one knows which VI must be inserted without the need to access Encyclopaedic information previously, for example (a move not allowed in DM). The phonological features may be very abstract but must be sufficiently specific at the same time to make the correct choice among roots at Spell-Out (Hagit Borer: p.c. shares a similar view, cf. Borer 2003, 2009[58]; but see Borer 1998: 174-176, 180-184 for some discussion about the need to insert phonological material from the beginning for both roots and functional heads).

Concerning the readjustment operations that some terminal nodes undergo on their way to PF (before they are provided with phonological expression), they are assumed to take place in the so-called Morphology component[59] and are necessary to explain mismatches between syntactic and morphological (PF) structure because the relation between morphemes (i.e. syntactic terminal nodes) and VIs is not always one-

---

[58] In her (2009) terms: "Roots merge as phonological indices. Phonological indices are exactly specific enough to ensure phonological faithfulness in the syntactic derivation, thereby excluding the derivation of *show* from *see* or an event-denoting noun such as *lesson* from some abstract non-existing verbal entry. A complete phonological matrix for a root is inserted on the basis of the index in the syntactic context created by the derivation".

[59] In Halle & Marantz (1993) readjustment operations are assumed to take place in a level called Morphological Structure. In later works (see e.g. Marantz 1997b), no level of Morphological Structure is assumed. Instead, morphology is regarded "as part of the Phonology, i.e., the interpretive component that relates an output of the computational system to PF (and here we'll assume that PF isn't Chomsky's "phonetic form" but some hierarchically organized prosodic structure) (p. 17)." However, note that, for example, the fusion, fission and impoverishment rules seem to have nothing to do with prosody and are not phonological in nature, but operate on syntactic nodes/features, in a way that is different from what phrasal syntactic operations do. That is, a morphological component seems to be needed anyway, despite DM's denial of it.

to-one (the default case).[60] Readjustment operations include fusion, fission, morphological merger and impoverishment, among others. Since the details of such processes will not be relevant for later discussion, only impoverishment and merger at PF will be briefly presented to illustrate the point.[61] Impoverishment, an operation originally formulated by Bonet (1991), consists in deleting morphosyntactic features from morphemes, so that a VI less specified than the VI requiring the deleted features can be inserted. There are several implementations of such an operation: Bonet (1991) treats it as feature delinking, while Noyer (1997) as restrictions on feature-cooccurrence. Regarding PF merger, DM makes use of such an operation to explain how the tense affix appears on the verb in English. Halle & Marantz (1993: 170) argue that the tense node lowers onto and merges with the main verb at MS [Morphological Structure] (e.g. *Dave and Sue **t** often danc–**ed** waltzes all night*). The operation of PF lowering can be seen as a weakening of the theory, since operations which are banned from taking place in syntax (e.g. there is a ban on downward movement in Narrow Syntax) are now allowed to occur at MS. It seems that the operations applying at this level are not restrictive enough, contrary to Embick & Noyer's (2007: 293) claim that PF processes are "modifications that are limited to minor operations that manipulate nodes in a sharply constrained fashion" (see the quotes in (62-65) below). PF lowering does not seem a minimal readjustment to us. In short, DM's claim that morphological structure is syntactic structure seems to be relaxed by PF operations.

In addition to readjustment operations taking place before spell-out, morphemes (i.e. nodes) and features can be added to the structure at this point as well. They are referred to as 'ornamental' morphology and are called disassociated morphemes/features: they satisfy language-particular requirements (e.g. case, passive) and are not crucial for semantic interpretation (see Embick & Noyer 2007: 305-309 for some examples and further discussion). Like the readjustment operations discussed above, the ability of morpheme/feature insertion after syntax seems a convenient solution to explain the extra morphemes/features which do not correspond to syntactic structure but are nevertheless present in the final outcome. If we stick to DM's basic claim that morphological structure is syntactic structure (in the default case), then one is

---

[60] See Aronoff (1994a) for an illustration of the complex mapping between morphological and syntactic structure.

[61] For the definition and exemplification of the other readjustment rules, the interested reader is referred to general works which give an overview of the framework and to more specific works in DM: e.g. Halle & Marantz (1993), Harley & Noyer (1999), Noyer (1997), and Embick & Noyer (2007).

forced to assume that the morphemes/features added at PF should also be present in the syntactic structure. However, this move is avoided in DM because it would greatly complicate the syntactic structure, an undesired result, and the PF solution is used instead.

Precisely because readjustment operations and morpheme/feature insertion processes are a weak point in DM (their claim that morphological structure is syntactic structure must be relaxed), DM proponents need to emphasize that both PF operations and disassociated morphemes/features are of a very restricted nature and are subject to language-particular constraints. Consider the following quotes which make the point clear (italics: ours).

(62) "… in DM the ordering, number, feature composition, and hierarchical positioning of terminal nodes may change in the derivation of MS [Morphological Structure], but only *in highly constrained and fairly well understood ways*".

Halle & Marantz (1993: 121)

(63) "In many languages –for example, Spanish, Russian, Latin, Latvian- word stems must have a Theme suffix, which has no syntactic or semantic role (…). It is natural to assume that such affixes are introduced by the rules that relate SS [Surface Structure] to MS. (…) like the Theme, Case and Agr morphemes are added to heads at MS in accordance with *language-particular requirements* about what constitutes a morphologically well formed word in that language".[62]

Halle & Marantz (1993: 135)

(64) "In the spirit of *restrictiveness*, then, one might suppose that the only features explicitly added in the "phonology" (…) are those that are completely absent from the syntax".
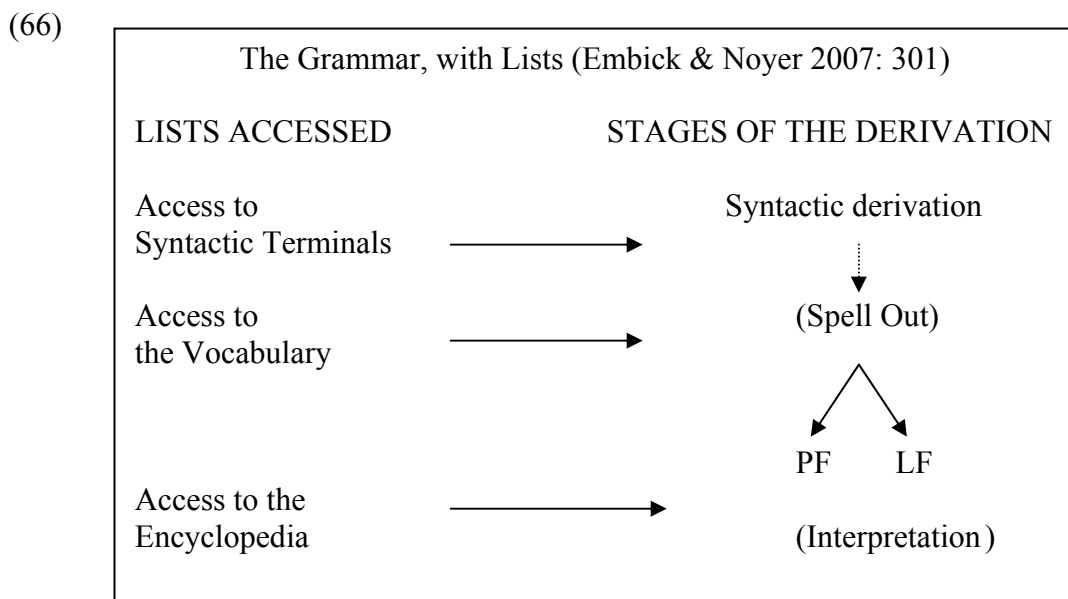
Marantz (1997b: 17-18)

---

[62] Chomsky's (1995a, b) Bare X-Bar Theory is still not incorporated in the early work of DM like Halle & Marantz (1993). See later works in DM (e.g. Marantz 1997a, b, 2001, 2007) for a full adoption of Chomsky's Bare Phrase Structure.

(65) "It must be stressed that the operations that apply at PF are *minimal readjustments, motivated by language-particular requirements*. Unlike the syntax, which is a generative system, *PF is an interpretive component, and the rules that alter the syntactic structures do not apply freely*".

Embick & Noyer (2007: 304-305)

To our mind, DM's claim that morphological structure at PF is syntactic structure (with one-to-one mapping) seems to be weakened by the existence of PF operations that can alter the initial syntactic structure and by the presence of disassociated morphemes/features. Despite our criticisms, one should be aware that such operations and additional material are not easily accounted for in any theory handling the same kinds of facts. Recall, though, that A&N's (2004) model handles these facts in the morphological component (cf. footnote 59).

In conclusion, this subsection has provided a brief overview of the framework of DM with special emphasis on those assumptions and claims that have some consequences for the architecture of grammar. The model of grammar assumed in DM is shown in (66).

(66)

The Grammar, with Lists (Embick & Noyer 2007: 301)

LISTS ACCESSED   STAGES OF THE DERIVATION

Access to   Syntactic derivation
Syntactic Terminals ⟶

Access to   (Spell Out)
the Vocabulary ⟶

                                           PF   LF

Access to the ⟶
Encyclopedia   (Interpretation)

1.4.2 Compounds in DM: Harley (2004, 2008b)

This subsection contains Harley's (2002, 2004, 2008b) assumptions concerning the nature of compounds (subsection 1.4.2.1), her analysis of what she calls synthetic modifier compounds (subsection 1.4.2.2) and synthetic argument compounds (subsection 1.4.2.3). Some problems regarding her analysis are pointed out, and an alternative account along the lines proposed by A&N (2004) is sketched.

1.4.2.1 Assumptions

Recall from the previous subsection that DM assumes that words are created by syntactic operations like head movement, as is shown below (italics: ours):

(67)   "Syntax, using conventional operations such as *head-movement*, plays a major role in constructing morphosyntactic structures, including *'word'-internal structure*"

Harley & Noyer (1999: 3)

(68)   "(…) *"words"* are assembled by *rules of the syntax*"

"Concerning the specific derivational mechanics at play in word formation broadly construed, we assume that in the normal case, complex heads are created by *the syntactic process of head movement*."

Embick & Noyer (2007: 290, 302)

Similarly, Harley (2008b: 1) derives compounds syntactically. More specifically, she treats them as incorporation structures à la Baker (1988).

(69)   "(…) compounds are incorporation structures, where non-head nouns incorporate into the acategorial root of the head noun, prior to its own incorporation into its category-defining n° head."

In earlier work (2002, 2004), though, Harley treats the mechanism of head movement as a conflation mechanism, a mechanism initially suggested and later rejected by H&K (2002). Harley adopts H&K's idea that conflation is associated with Merge. H&K understand conflation as copying the phonological material, the "p-sig" in their terms,

of the sister head into the higher phonological empty head, i.e. a head with a defective p-sig. Harley (2002, 2004) represents the defective p-sig by the feature ±affix,[63] which can be generated on any head. Harley (2004: 5) provides her mechanism with a definition:

(70)   "… only the p-sig of the label of its sister may be conflated during merge of a [+affix] head. The p-sig in the label of the sister is a copy of the p-sig of the *head* of the sister. Any p-sigs *within* that constituent will not be eligible for conflation, unless they have previously been conflated into the label of the head of the sister".

Although Harley (2008b) is not really clear about the nature of head movement[64], we could view the process as being phonological and syntactic at the same time: phonological in the sense described in (70) and syntactic in the sense that it occurs during the syntactic derivation (Narrow Syntax) and is framed within DM. Such an understanding of head movement is compatible with Harley's analyses of compounds (see the following two subsections), from which it will also be clear that internal arguments and modifiers of roots are merged with roots first, before categorization of the root. In other words, both synthetic modifier compounds (e.g. *quick-acting*) and synthetic argument compounds (e.g. *truck driver*) are given the same analysis.[65]

1.4.2.2 Synthetic modifier compounds

Harley (2008b) defines synthetic modifier compounds like *quick-acting* as those compounds in which the first element (on the surface; *quick*) acts as the modifier of the second element, or rather of the root contained in the second element ($\sqrt{ACT}$). Harley provides the tree in (71b) for *quick-acting*.

(71)   a. quick-acting baking powder (It acts quick(**ly**))

---

[63] Following Harley (2002, 2004), the label ±affix does not necessarily imply that the heads are necessarily morphophonologically affixal in nature, though they are most of the time.

[64] That Harley (2008b: 9, fn. 3) does not find crucial the technical implementation of head movement is shown in: "The mechanism of head movement could be either the conflation mechanism adopted in Harley 2004 or the phrasal-adjunction-plus-morphological-merger mechanism proposed in Matushansky 2006."

[65] The terminology used to refer to the different types of compounds discussed here (e.g. synthetic modifier compounds and synthetic argument compounds) is taken from Harley (2008b). See chapter 2 for discussion on the different terms used in the literature and for my own choice.

b.

```
                              aP
                        ┌──────┴──────┐
                       a°             √P
                   ┌────┴────┐     ┌───┴───┐
                   √        a°   √ACT      aP
               ┌───┴───┐     │     │       △
              a°     √ACT   -ing   aᵉt    quick
           ┌───┴──┐    │
         √QUICK  a°   act
           │      │
         quick    Ø
```
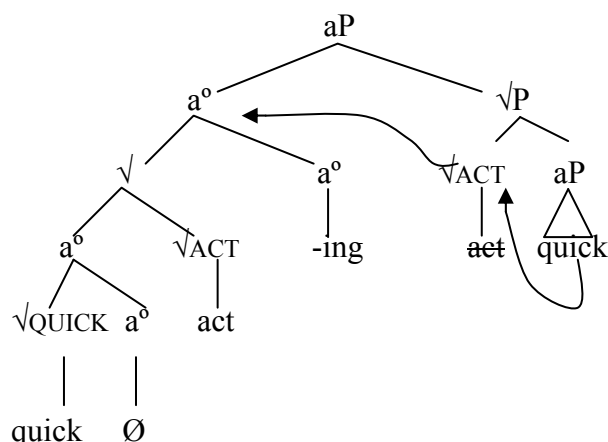
As noted in the previous subsection and illustrated in (71b), the adjectival modifier *quick* merges with and incorporates/conflates into the root $\sqrt{ACT}$ before the root undergoes categorization. Afterwards the complex *quick-act* incorporates/conflates into the adjectivalizing head *–ing*. This analysis raises a number of questions.

First, if compounding is really the result of incorporation/conflation based on the syntactic counterpart of a compound, then there is no reason why *quickly* as such (compare (71a) with (71b)) cannot occur inside the compound (which is what one would expect to occur), unless one resorts to some convenient readjustment rules which delete the final suffix on the way to PF. Even though it has long been argued that the so-called adverbs are a derived category (e.g. Emonds 1976: chapter 5 and Radford 1988: 137-141 for English, Bartra & Suñer 1992 for Romance, Baker 2003: 230-237 cross-linguistically) and, consequently, the *–ly* suffix could be a reflection of the adjective (the base form from which the *ly*-form would be derived) appearing in syntax, Harley claims that the compound is formed on the basis of the syntactic structure (the *–ly* suffix is present there) and hence she has to account for the omission of *–ly* in *quick-acting* in some way.

Second, it is not clear to us why *quick* has to incorporate if a non-incorporation syntactic structure is available. To account for the coexistence of incorporation and non-incorporation structures, Harley (2002, 2004, 2008b) makes use of a [+/- affix] feature present in the root. Recall from (70) that if the root is provided with a [+affix] feature, it means it is defective and needs phonological copying of the sister head, the result being a conflated structure (*quick-acting*). By contrast, if the root is not defective, i.e. is provided with a [-affix], then no conflation process takes place and the result is an

analytic structure (*act quickly*). As will be discussed in the next subsection, we find this [+/-affix] feature a bit ad hoc and it reminds us of the look-ahead problem: depending on the structure one wants to generate, the affix will have one or the other value. Maybe such a proposal would make more sense in the case where roots are really morphologically bound (e.g. *vis-* in *visible*), but not in the case where they can freely form an independent word in syntax (e.g. *act*).

Third, more problematic is the fact that the analysis in (71b) does not capture the fact that *-ing* only attaches to verbs (already categorized roots; see e.g. Grimshaw 1990, Marantz 1997a but see Borer 2009 for a different view).[66] Harley's (2008b) preliminary solution is to add a categorizing vº above the √P once *quick* has incorporated into √ACT. Given that this analysis predicts the grammaticality of verbal compounds and, in fact, compounds like **to quick-act*, **to meat-eat*, **to corn-fertilize* and **to truck-drive* are all ungrammatical, Harley then proposes that in languages like English (but not Mohawk, for example) the head vº is prohibited from hosting compounding: it cannot contain more than one root in its base position.[67] Her proposal, though, is not supported by compound verbs like *to computer-generate*, *to Chomsky-adjoin*, *to steam-clean*, *to deep-fry*, among many others (see chapter 2). Although some of these NV compounds may come from backformations, we consider that they all have the same status as non-backformed compounds. When children learn these compounds, they do not know whether they are base-generated or whether they come from backformations (see McIntyre 2009 for related discussion).

An alternative analysis was suggested in subsection 1.3.1. If we assume a model of grammar, along the lines proposed by A&N (2003, 2004, 2007), there are two generative systems (one for words and one for phrases) which compete with each other for the combination of lexical categories (there are no acategorial roots, cf. Don 1993). Recall that in non-polysynthetic languages like English, the syntactic merger wins over the morphological merger iff the two mergers have the same semantics and merger of

---

[66] Revealingly, some nonsense words in Catalan and Spanish (imitating the English construction of V+ing for the creation of sports like *swimming*, *biking* and *hiking*) also seem to be subject to this constraint: Cat. *bicing* (bic(i)+ing bike+ing, in the sense of 'to bike somewhere'), *panxing* (panx(a)+ing belly+ing, 'to laze around'), *sofing* (sof(à)+ing sofa+ing, in the sense of practising the sport of being on the sofa, i.e. 'to relax'). Although the base word to which –ing is attached does not exist as a verb in Catalan, it acts as if it were one.

[67] Harley (2004: 13, fn. 13) discards the possibility that object incorporation into English vº is not allowed for case reasons since compounds with unergative and unaccusative verbs are ungrammatical: e.g. **The snow fast-fell* (vs. *fast-falling snow*).

categories, two conditions which are summarized in the constraint given in (26), repeated below for convenience.

(26)   Let α_1 and α_2 be syntactic representations headed by α. α_1 blocks α_2 iff
         (i) in α_1 (a projection of) α is merged with (a projection of) β in syntax, while in
             α_2 (a projection of) α is merged with (a projection of) β in morphology, and
         (ii) the semantic relation between α and β is identical in α_1 and α_2.

By contrast, morphological merger is possible iff there is no syntactic competitor: when different categories merge in morphology and in syntax (e.g. if a category-changing suffix is present, morphological merger is required) or when the syntactic and morphological structures involve different semantics. Such a morphosyntactic competition model can explain the following contrasts:

(72)   a. √to act quickly
         b. to quickly-act

(73)   a. √to eat meat
         b. to meat-eat

(74)   a. to computer-generate
         b. to Chomsky-adjoin

The examples in (72) and (73) illustrate the same point. Regarding (72), an adverb merges with a verb in syntax and in morphology with the same semantics in the two mergers. Because of the morphosyntactic competition, the merger of an adverb and a verb takes place in syntax and *to act quickly* wins over **to quickly-act*. The ungrammaticality of compounds like **to meat-eat* and **to truck-drive* (73b) receive a similar explanation. Given that the verb and the noun can also merge in syntax with the same semantics (e.g. *to eat meat*; *meat* is the internal argument of the verb), the syntactic merger wins over the morphological merger. As noted above, similar compounds are grammatical, though: e.g. *to computer-generate* (74). Despite apparent similarity (i.e. a noun merges with a verb), the semantics of the morphological merger is not identical to that of the syntactic merger: in the compound, the computer is the

instrument (means) by which the generation of something is carried out, whereas in the syntactic merger (i.e. to generate a computer) the computer is the internal argument of the verb. In addition, if the semantics of the compound wants to be preserved in the syntax, then additional lexical categories are necessary (e.g. to generate by means of a computer), with the consequence that the mergers in morphology and in syntax are no longer identical and competition is suspended.

In short, Harley's (2008b) analysis of modificational synthetic compounds should tackle some thorny questions which it faces at the moment. By contrast, the competition model (A&N 2004) seems to fare much better with those cases which proved difficult in Harley's DM-based account. The following subsection is devoted to Harley's analysis of synthetic argument compounds.

1.4.2.3 Synthetic argument compounds

Harley (2008b) presents the English *one*-replacement contrast in (75), from which she concludes that apparent arguments of nouns (*student of chemistry*) are in fact arguments of the underlying root in the noun (√STUD, which can also explain *to study chemistry*). By contrast, similar structures in which a noun is followed by an adjunct PP behave differently. In this case, the PP merges with an already categorized root. Based on the paradigm established in (75), Harley provides the analyses in (76) and argues that anaphoric *one* takes an nP as its antecedent. Only on this assumption does it follow that the argument is merged inside nP, hence is selected by the uncategorized root. On the (usual) assumption that *one* takes an N' as its antecedent, the conclusion is that *chemistry* is an argument of the noun in (75a).

(75)   a. ?*The student of chemistry and this one of physics sit together

      b. That student with short hair and this one with long hair sit together

(76)
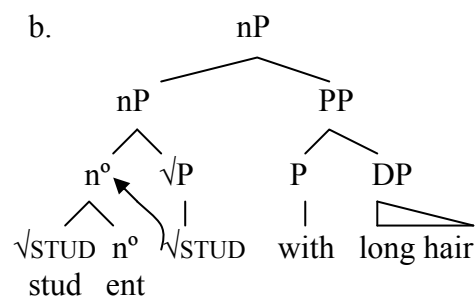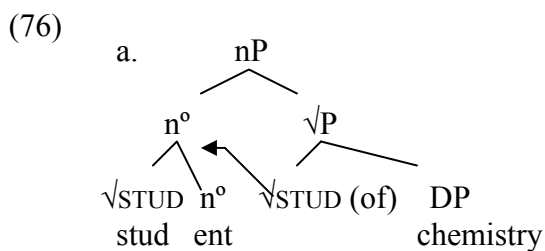
Harley's (2008b: 9) statement that "(internal) argument selection is a property of roots" is seen as positive because roots contain the encyclopaedic information which can tell whether an internal argument is needed. However, Harley's view is incompatible with the DM's claim that accessing encyclopaedic information requires that a phase is sent off to PF and LF, and that phases are established after a root is categorized (once the root is merged with a categorizing head, Marantz 2001, 2007). Hence, an unsolvable conflict, as far as we can see. In addition, if "(internal) argument selection is a property of roots", then it is difficult to explain why agent-denoting nominals like *student* and *driver* (based on the roots √STUD and √DRIVE) need not appear with their internal argument (vs. e.g. *chemistry student, student of chemistry*).[68] Also, whether or not the complement needs to be introduced by a dummy case-assigning preposition like *of* or not depends on whether the higher categorizing head turns out to be v° or n° (or a°), which is difficult to combine with bottom-up derivations that are restricted by something like Chomsky's (1995a) Inclusiveness condition. In general, within 'proper' (phrasal) syntax there do not seem to be cases where a higher head can have such non-local influence on the shape or other properties of a complement of a lower head.

Harley's statement that roots select (internal) arguments is further questioned by idiomatic readings: such readings show that the root must already be a verb when it merges with its internal argument (77a). The idiomatic reading is lost under inheritance (when the internal argument merges with a derived noun, and not with the underlying verb in the derived noun; 77b). In short, the availability of having or not having an idiomatic reading does not seem to be related to the properties of the root (the same root is present in both cases). Rather, it seems that the idiomatic reading can only be preserved if the noun and verb merge directly, as in (77a). Let us now consider the internal structure of *troublemaker* in (77c).

---

[68] Harley's (p.c.) thoughts on this puzzle are as follows:

(i) "The long answer (begin here my current speculations) has to do with the famous fact discovered by Grimshaw: the presence of an internal argument in nominals is sensitive to the presence or absence of an event interpretation, first identified by Grimshaw in distinguishing between deverbal event nominals and superficially identical result nominals, of course -- the former require the internal argument, the latter forbid it.
These examples with 'student' of course are not event nominals but agent-denoting nominals, but I think the same distinction perhaps applies; 'a student of chemistry' could include the implication that the person actually has engaged in chemistry-studying events, while 'a student' might not imply any events of studying have actually taken place. This would require a lot of pretty subtle investigation... teasing out distinctions between things like 'a frequent student of the classics' vs 'a frequent student', if there are any to tease out... (…)".

(77) a. John always makes trouble.

  b. #John is a maker of trouble.

  c. John is a real troublemaker.

Given that (77c) is grammatical and has an idiomatic reading, we can conclude that the noun and verb have merged directly, i.e. *to troublemake* (see A&N 2004: 54-59 for this example and for further data), with the consequence that the root is already categorized as a verb before it merges with its internal argument (contra Harley 2008b).

In addition, if internal arguments are a property of roots, then the source of the internal arguments of verbs like *industrialize* and *legalize* is not clear to us. For example, the underlying root in *industrialize* would probably be √INDUSTRY. Such a root does not require any internal argument, and yet the presence of an internal argument is necessary in *industrialize*, once the root has been categorized as a verb (*They industrialize *(the city)*). In conclusion, it is not the properties of the root that explain the presence or absence of an internal argument, but rather the extra material added to the root (Mateu 2005: 227-229 reaches the same conclusion independently: in a nutshell, his theory of impossible primitives predicts that a root is always associated with a non-relational element, hence the impossibility of taking a complement. Relational elements, by contrast, do take complements). In addition, if internal arguments were a property of roots, as Harley claims, words like *transportation* and *transportal* should have the same argument structure because they are based on the same root, but grammaticality judgments point to the opposite direction: *public transportation* vs. **public transportal* (cf. see Borer 2009 on this point).

Despite the problems just discussed concerning the alleged selection of internal arguments by roots, let us consider how Harley (2008b) analyses synthetic argument compounds by her applying to them the structure of (76a). Accordingly, compounds like *script-writer*, *truck-driver*, *drug-pusher*, *car-chasing (dog)* and *grass-clipping (machine)* have the structure depicted in (78). The root plus the internal argument can be nominalized or adjectivalized (if, for example, the category-changing suffix is *–er* or *–ing*, respectively). For concreteness' sake, the compound *script-writer* is illustrated below (cf. Harley 2008b: 11).

(78)

nP
  n°   √P
 √P   n°   √WRITE   nP
nP √WRITE -er write   n √SCRIPT
√SCRIPT n write   Ø script
script Ø
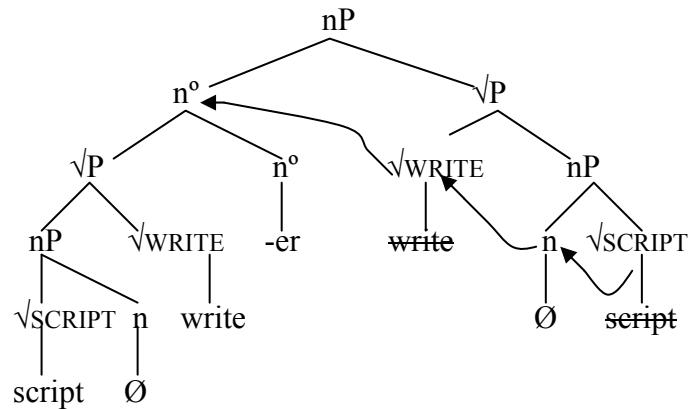
As already noted in the previous subsection for modificational synthetic compounds (*quick-acting*), Harley (2002, 2004) accounts for the coexistence of compounds (*scriptwriter*) and their analytic counterpart (*writer of scripts*) by means of a [+/- affix] feature. That is, the root √WRITE is provided with a [+/- affix] in the numeration. Depending on the value, there will or will not be an incorporation structure (cf. H&K's 2002 conflation mechanism). In other words, there are the following two possibilities:

(79)   a) If the root has a plus value, it means it is defective and needs phonological material from the sister head down below. As a result, an incorporated structure is derived.

b) If the root has a minus value, it means it is not defective and no copying of phonological material from the sister head is needed. As a result, a non-incorporated structure is derived.

As was hinted at in the previous subsection, we are a bit sceptical about the [+/- affix] feature. If the very same root is present in both structures, having a [+/- affix] seems an easy solution to get the right structures. How do we know which value [+/- affix] gets inserted in the numeration? How can we distinguish when one word like *write* has one feature or another? The look-ahead problem seems to be present: if one wants to derive a compound, then one has to assume that the root is provided with a [+affix] feature in the numeration; by contrast, if one wants to derive a phrase, then one needs to assume that the root is equipped with the [−affix] feature in the numeration.

In Harley (2008b), the [+/- affix] feature is only used to account for modificational synthetic compounds (e.g. *quick-acting*). In the case of argumental synthetic compounds, internal argument nouns incorporate for case reasons (e.g. *script* in *scriptwriter*, *truck* in *truck-driver*). For cases like *trucks-driver and **[the truck]-driver* in which the internal argument cannot incorporate, Harley argues that DPs are prevented from incorporating. Her argumentation is as follows: if an nP merges with number or determiner material (as is the case in the previous examples), the case feature of the nP is checked DP-internally, and the case feature is no longer available for incorporation.[69] However, it is not clear to us how case can be checked DP-internally in the case of *drive trucks* and not in the case of *truck-driver*, given that the same LI *drive* is present in both structures. Even though there is more functional material above *trucks* (e.g. Numº), we do not see why in a syntactic analysis of compounds *trucks* cannot undergo head-to-head movement via the functional heads until it reaches the root √DRIVE.[70] Consequently, we regard Harley's solution as a little ad hoc, which also seems to be designed to explain the fact that modifiers in general (e.g. adjectives, PPs) can be neither incorporated nor stranded (see, e.g., A&N 2004, 2007, Padrosa-Trias 2007a for the discussion of problems like stranding and undergeneration by the movement account, among others, involved in a syntactic movement analysis).

In addition, given that compounds (e.g. *truck-driver*) have a generic/habitual reading, one expects that the underlying syntactic structure from which they are derived

---

[69] Note that Harley's account cannot explain the contrast between (ii) and (iii) below, given that they are based on the same underlying structure: √*AMUSE/FRIGHTEN the child*, and accordingly the same behaviour is expected: the internal argument should not be able to incorporate in either case because it is introduced by a determiner, which prevents incorporation for case reasons. One could try to explain the contrast by appealing to Marantz's (2001) distinction between root nominalization in the cases of (ii) (e.g. *child-amuser*), with no v present and hence no possibility of projecting an agent, and verb adjectivalization in the cases of (iii) (e.g. *child-amusing*), with the presence of v and the projection of an agent. Recall that –*ing* attaches only to verbs (see Mateu 2009 for related discussion). This initial solution is still not enough, though, because the domain where the root merges with its internal argument is shared in the nominal and adjectival structures, and hence cannot explain why incorporation is possible in one case but not in the other case.

(i)   T.V. frightened/amused the child.
(ii)   *A child-frightener/child-amuser.
(iii)   A child-frightening ride, the child-amusing clown

Harley (2004: 11-12, fn. 12)

[70] Under Minimalist assumptions (Chomsky 2000, 2001, 2004, 2005, 2007), Harley's explanation cannot hold. Independently of whether the object is plural or is preceded by a determiner, the element responsible for assigning case to *truck* (i.e. giving it a value) is the verb *drive*. Following Chomsky, the nominal (internal argument) values the φ-features of v*/V, which in turn assigns accusative case to the object. On Harley's account, though, the head valuing the case feature of the object seems to change in an ad hoc manner to account for the data (the verb in compounds and some functional head in non-incorporation structures).

should have the internal argument in the plural form to account for the correct reading (e.g. *to drive trucks*).[71] If this is the case, then the Num° node should also prevent incorporation in this case (according to Harley, case would be checked DP internally) and yet the result is a compound, a difficult situation as the theory stands. One more prediction of Harley's account is that if only bare objects can incorporate (for case reasons), incorporation should only be possible with objects like mass nouns, which lack any visible number and determiner material. However, as has been shown, these are not the only cases in which incorporation is allowed: in addition to compounds with an incorporated mass noun like *paper shredder*, *glass maker*, *cement mixer* and *gold digger*, there are compounds like e.g. *truck-driver*, *bicycle-repairer*, *dish-washer*, *bookseller*, *task assignment* and many others (see chapter 2 for more examples).

Also problematic for Harley's account are examples like those in (80).[72]

(80) a. drug-pusher to children
      b. truck-driving across the country
      c. horse-jumping over fences
      d. book-giving to children
      e. gun supplier to the army

Concerning the examples in (80), Harley's (2002, 2004, 2008b) incorporation/conflation analysis predicts their ungrammaticality. Let us consider why. Harley assumes a SC analysis (cf. Hoekstra & Mulder 1990) for the underlying structure from which the incorporated expressions in (80) are derived. For example, in (80a), the root √PUSH takes a PP as its complement, with the P *to* as the head, *drugs* in the specifier position and *children* in the complement position. Harley follows a strict incorporation/conflation analysis and the prediction is that only elements occupying the sister head position can incorporate and that no element from the specifier position can incorporate (see Harley's statement in (70) and Baker 1988 for further discussion). Hence the incorporation of *drugs* is an illicit move in an incorporation/conflation analysis and should be ungrammatical. This is in agreement with Harley's judgments of (80a-c) (see also Selkirk 1982: 37), but this is certainly not the case with all speakers.

---

[71] Even if one derives *truck-driver* from 'to drive a truck' (as in *He drives a truck for a living*), the same problem is still present: *truck* is not prevented from incorporating despite the presence of a determiner.
[72] Some of the examples in (80) are borrowed form Selkirk (1982: 37, ex. 2.40).

The informants consulted regard all structures in (80) as fine (contra Harley; for a different analysis see McIntyre 2004).[73]

In addition, as the theory stands, nothing prevents the generation of *to-pusher drugs children* since *to* is the head of the PP, the sister head of the root √PUSH. To avoid such an incorporation/conflation structure, Harley gives different possible solutions. First, there may be "null prepositions with defective p-sigs in English" (Harley 2002), a position she adopts to explain adjectival passive compounds like *expert-tested*, *pan-fried* and *snow-covered* (examples from Roper & Siegel 1978: 242; see Padrosa-Trias 2007a for a critical review of Harley's analysis). Second, Harley entertains the possibility of having the preposition adjacent to its object in syntax for case reasons, an option which is disregarded in favour of her third possibility: the prepositions present in syntax (e.g. *to*, *across*) do not have a [+affix] specification. That is, according to Harley, some morphemes may be morphologically specified as free or bound: roots may vary in their affixal specification but abstract morphemes may not (see the Affixal Determinism principle in Harley 2004: 11).

Also, note that Harley's initial insistence on the separation between roots taking internal arguments directly and already categorized roots taking adjuncts (cf. 76) vanishes once compounds other than argumental synthetic compounds (e.g. *truck-driver*) are considered. Recall that modificational synthetic compounds (cf. *quick-acting* in (71)) were derived in the same way as synthetic argument compounds, and so are primary root compounds like *nurse shoes* and *alligator shoes* (see Harley 2008b for a representation of such compounds). As a result, compounds like *truck-driver* and *home-making* in which *truck* and *home* can be interpreted either as the internal argument of the underlying verb, or as an adjunct of the derived noun, will be given the same analysis. In short, all compounds seem to be derived in the same way and Harley's argument to merge the root with the internal argument first is undermined: there seems to be no strong reason for not categorizing the root before it merges with the internal argument (or an adjunct for that matter). Finally, note that Harley (2008b: 17) leaves it "up to the interpretive component to construct some plausible relationship between the incorporated noun and the head noun", which she applies to primary root compounds but, in fact, it also seems applicable to all compound types.

---

[73] Unexpectedly, though, speakers did not accept similar conflated/incorporated structures: e.g. *toy-handing to babies* and *boot-putting on the table*.

In short, Harley's (2008b) claim that internal arguments merge with the roots that select them before the roots are categorized has been questioned. Some arguments have been provided that show that the root must already be categorized before it merges with its internal argument (recall, for example, the discussion around *industrialize*). Harley's analysis of argumental synthetic compounds has also been called into question, in particular the reason for having incorporation vs. non-incorporation structures (the case feature). Also, some data have proved difficult to tackle in Harley's proposal (e.g. 80). Finally, all types of compounds receive the same analysis, which further questions Harley's initial distinction between internal arguments and adjuncts (merger with roots vs. already categorized roots).

The next subsection contains some questions which we think are either controversial in the DM framework or are still open questions to be answered by future research.

1.4.3 Discussion: debatable questions in DM

This subsection presents five main questions (five claims made in DM) which we think merit further study due to being either controversial issues in DM or because it is still too early to state conclusive claims about them.[74]

First, the alleged existence of some roots in DM seems to be questionable. For example, words like *visible*$_{a^{\circ}}$ and *vision*$_{n^{\circ}}$ are assumed to have the following internal structure: $[[\sqrt{VIS}]_{\sqrt{}} a^{\circ}]_{aP}$ and $[[\sqrt{VIS}]_{\sqrt{}} n^{\circ}]_{nP}$ respectively, with the root $\sqrt{VIS}$ and with *–ible* and *–ion* as the category-creating terminal nodes. Similarly, words like *pomposity* and *pompous* are assumed to contain the root $\sqrt{POMP}$, and words like *porosity* and *porous* the root $\sqrt{PORE}$ (see Marantz 2001 for such data).[75] For most speakers, this is clearly not the case: they cannot identify the roots on which the derived words are based. Other words which are not decomposable for some speakers include *revolution*, *residence*, *permutation* and *activities*, due to the idiosyncratic relation between the base verb and the derived noun (see Chomsky 1970 for other examples).

Also, DM treats all roots in the same way, which we view as potentially problematic. We think a distinction should be drawn between roots which are

---

[74] For a critical view of DM, see e.g. Williams (2007).
[75] Following the same pattern, words like *altitude* and *attitude* will presumably be analysed into the categorizing suffix *–tude* (not a morpheme!), with the remaining part being the root (cf. Williams 1994).

semantically transparent and those which are not, because only the former can be used by speakers to create new forms. That is, Marantz's (2001, 2007) account of roots may be valid but only for semantically transparent roots to which speakers can apply word-formation processes and, for example, it may not be applicable to classical-based roots of which most speakers are not aware of their form or meaning, and are then unable to create new words. To illustrate the point, the Catalan forms with the root √SCRIURE (in words like *prescriure* 'prescribe', *inscriure* 'inscribe, enroll', etc.) all come from Latin (e.g. *inscriure* comes from *inscribĕre*, dating back to 1839)[76] and are opaque to most speakers (probably non-linguist speakers and those speakers who have not received a formal education in classical languages), with the result that they are unable to create new words based on the Catalan √SCRIURE (Latin *scribĕre*).

The division between semantically transparent and semantically non-transparent roots seems to roughly correspond to the division between "outer/non-first phase" and "inner/first phase" morphemes of Marantz (2001, 2007). For us the properties typically associated with each type of morpheme are not surprising (see Marantz 2001: 14-15 for the lists of properties). Unlike outer/non-first phase morphemes, inner/first-phase morphemes tend to not be decomposable by everyday speakers and, together with the root, are seen as a unit, as a whole. In most cases, the properties associated with the two types of morphemes seem to follow from the history of the language (e.g. Latin- and Greek-based roots together with the first phase/inner morphemes tend to be seen as a whole by speakers).

One might argue that, despite not being perceived by most speakers as building blocks in the language, inner/first phase morphemes still determine some specific behaviour of the words they are part of in the syntax. For example, Marantz (2001: 21; see also Marantz 2003) claims that prefixed verbs like *destroy* are all predicted to be transitive due to the presence of the prefix: *destroy* is decomposed into the root √STROY and a SC complement, of which the prefix *de-* acts as the predicate and thus necessarily requires the presence of an inner subject (i.e. the internal argument of *destroy*), as in *destroy *(the city)* (see also Harley 2007 who reaches the same conclusion independently). Whereas the prefix *de-* may have been the source for having a transitive verb in e.g. *destroy* and Cat. *decidir* 'to decide' originally, languages evolve and the prefix may now not have the function it used to have. In some cases, the prefix has

---

[76] Source: *Gran Diccionari de la llengua catalana* (GDLC)

become opaque and indistinguishable from the (alleged) root, i.e. the pattern may have become a fossil of a rule that was active some time ago. This seems to be the case for the Cat. verb *decidir*, which does not require the presence of an internal argument and a sentence like *Els estudiants decidim* 'The students decide' is perfect (as is its English counterpart, in fact) (Other prefixes behave in the same way, e.g. *ob-* in Cat. *obstruir* 'to obstruct'). In short, the conclusion seems to be that decomposing words too much may not give the right results, which can be a consequence of roots becoming non-transparent semantically and blurring the original pattern underlying the word. This conclusion will lead us to the last question concerning roots: the difference between transparent and non-transparent roots and its effects in relation to the lexicalist hypothesis that word-sized units are a special unit in the grammar.

Marantz (1997a: 205-213) criticizes the lexicalist claim that the phonological word has some special status in the grammar by being associated with special prosodic structure, meaning and structure/meaning correspondences. Marantz gives arguments against such associations. For example, he observes that syntactic and prosodic structure is not isomorphic at any level (the phonological word included, cf. Jackendoff 1997, 2002, see section 1.1). He further notes that special meanings may not be uniquely tied to words since units smaller (see the discussion on stative and agentive passives in Chichewa which are expressed by means of suffixes in Dubinsky & Simango 1996) and bigger (e.g. 'light verb' constructions, idiomatic phrases) than words can also have special meanings (see also Marantz 1997b and 2001 for further discussion).

We agree with Marantz on his criticisms against the word being a special domain in the grammar in general. Recall that words and phrases are treated equally in A&N's (2004, 2007) theory, a position we adopt. For us both words and phrases form part of a generative component, although a different one in each case, morphology and syntax (see subsection 1.2.1 for evidence for the two generative systems). For us both words and phrases can be built compositionally, and idiosyncrasy can equally apply to both objects: there are both idiosyncratic words (native words included) and phrases (e.g. idioms). On the other hand, we disagree with Marantz (1997a: 212-213) when he says that what is special is the root: "Things with special meaning are roots". Recall the discussion above concerning the distinction between semantically transparent and non-transparent roots. In short, word-sized units can have a special sound, meaning and structure/meaning correspondences if they are based on a non-transparent root or the root is indistinguishable from the word (e.g. English *cat*, Catalan *gat* 'cat').

The second question which will be discussed in this subsection is a basic claim in DM: the claim that roots are acategorial. At first sight, a good point about having acategorial roots is that of economy. The same root can belong to more than one category depending on the context in which it occurs (e.g. *walk*$_{N/V}$). However, all regular and productive affixes (i.e. non-first-phase heads) are claimed not to attach to roots but to a head that has already attached to the root and given it a category (n°, v° or a°). This is the case of the deverbal suffix *-er* in English, for example: this affix attaches to verbs to create nouns (e.g. *painter*, *driver*, *dancer*; cf. Lieber 1992: 54). Having acategorial roots is then irrelevant for phases higher than the first one. For cases like the suffix *–er* it seems that the most economical option would be not to go through a verbalizing functional head but to simply attach to a verbal base (e.g. *drive*).[77] The advantages of having acategorial roots are then reduced to those affixes that attach to roots directly and even some of these cases seem doubtful: the suffix *–ous* is claimed to attach to roots but some of them could be treated as already categorised as nouns. Such is the case of *virtuous < virtue, glorious < glory*. In short, the economy argument favouring acategorial roots is greatly weakened (if not eliminated entirely).[78]

Still in relation to acategorial roots, a question that springs to mind is why nouns like *cat* should come from an acategorial root if it is always a noun. A plausible answer could be that the most economical alternative seems that *cat* is a noun right from the beginning with no need to categorize it. However, Marantz's (2001: 12) answer is as follows: "'cat' as a verb has no obvious meaning/use, although it can be given fine meanings contextually ('Meowing and scratching in imitation of his pet feline, Fred

---

[77] The suffix *–able* behaves in the same way as *–er* in the sense that it suffixes only to verbal bases to produce, in this case, adjectives (e.g. *wash*$_V$-*able*$_A$; cf. Lieber 1992: 54). Similarly, the Spanish suffixes *–(i/e)dad* and *–mento* only attach to adjectives to create abstract nouns and adverbs, respectively (e.g. *igual*$_A$-*dad*$_N$ 'equality', *necia*$_A$-*mente*$_{Adv}$ 'stupidly', cf. Varela 1999: 273).

[78] See e.g. Kayne (2009) who suggests that a category-creating *n* may not be needed. In addition, note that in DM derivational suffixes like *–ous* (as in *virtuous*) and *–ety* (as in *variety*) are considered functional categories which categorize the roots to which they are attached, not an uncontroversial claim. Finally, consider Baker's (2003: 266, fn. 1) quotation, which suggests a parametric difference between having or not having categorial roots cross-linguistically:

(i) "I strongly suspect that the freedom of roots to switch categories is much freer in English (and languages like Tongan, Mandarin, and Hebrew) than it is in languages like Mohawk, Edo, Chichewa, and Australian languages. This could raise questions about the suitability of the Marantz/Borer theory of category-neutral lexical heads. At least the implications of such a "parameter" of variation for this view have not been considered."

catted around the house for hours')." We will leave it up to the reader to ponder which answer is the most appropriate one.

The issue of having (a)categorial roots is being hotly debated in the literature at the moment and we do not intend to resolve it here, which explains why it has only been mentioned briefly. The interested reader can consult works that address the topic more thoroughly: e.g. Baker (2003), Bauer & Valera (2005), Borer (2003, 2009), Brattico (2005), Don (1993, 2003, 2004), Don, Trommelen & Zonneveld (2000), Lieber (2006), McIntyre (2009) and Marantz's work in DM to name but a few references.

A third question that will also be briefly touched upon is concerned with the different flavours associated with the category-creating terminal nodes, which can be seen as a bit ad hoc. Harley (2008b) associates the aº head with the following flavours: 'characterized by' (as in *careful*, *comfortable*), 'able to be' (as in *edible*) and 'like' (as in *yellowish*, *boxy*). These characterizations seem to be replete with idiosyncratic semantics, somehow responding to the needs of each affix, and bring us to the distinction drawn by Mateu (2002) in (15), repeated below for convenience.

(15)   "Meaning is a function of both (non-syntactically transparent) *conceptual content* and (syntactically transparent) *semantic construal*."

Following (15), what is syntactically relevant for meaning is the semantic construal, not the conceptual content which is full of idiosyncrasies. Although Harley's (2008b) work is a syntacticocentric approach to word-formation (similar to Mateu 2002 in this respect), the distinction between conceptual content and semantic construal is blurred and is not maintained, as can be seen from her labels which denote conceptual semantics, an incongruity in the system.

More plausible flavours for the aº head should get closer to the nature of the flavours which have already been proposed for the vº head. For example, Marantz (2001: 21) considers functional category verbs like 'be', 'have', 'do', which are more in line with the semantic construal part of the definition of meaning in (15). Similar flavours have been proposed in other work (see, e.g., Jackendoff 1990, Harley & Noyer 2000, H&K 2002, Mateu 2002, Baker 2003, Harley 2008b).[79] The exact flavours which

---

[79] For example, Mateu's work (e.g. Mateu 1999: 4-5, Mateu & Amadas 2001: 9-10, a.o.) defends the view that relational heads (not only verbs) are assigned lexical semantic properties (in a binary fashion) like

should be associated with the aº head is still an open question, which needs to be further investigated.

The fourth question that will be mentioned in this subsection has to do with another claim made in DM, the claim that there is competition for insertion between VIs like, for instance, –ness and –ity in the nº head, with the flavour 'the property of' in words like *happiness* and *elasticity* (Harley 2008b: 6). Marantz (2001: 11) makes the same point, which is illustrated in (81) (see Embick & Noyer 2007: 298-299 for similar examples and related discussion).

(81) "Oxen, *oxes: –en competes with –z for insertion into a [plural] node, and wins out here since it is specified to attach to 'ox'
Reversibility, *reversibleness: -*ity* competes with –*ness* for insertion into a N-forming node that merges with adjectives. –*ity* wins out here since it is specified to attach to –*able*"

According to the DM account, competition between VIs is resolved when the material to which a particular VI is attached is examined. For example, one knows that the VI –*en* (and not –*s*) is chosen to indicate plurality once you know that the root is *ox*. Similarly, the nº head will be realized as –*ness* or –*ity* subject to the root the nº has merged with (*happy* vs. *elastic*). This account of competition reminds us of listing. If one has to specify each root and morpheme that is able to attach to, say, -*ity*, this view is very similar to one having a Lexicon (as it was used in lexicalist theories).[80] Our claim is that the choice of one or other VI is not a matter of competition: VIs are simply subject to different selectional requirements, and consequently are merged with different types of bases (e.g. native vs. Latinate).

In addition, note that according to Marantz (2001: 11), one should be asking about the productivity and distribution of morphemes: "(…) asking about the distribution of N(-creating) nodes merging with roots or merging outside V nodes (…)", instead of asking about the distribution of specific VIs like –*ity*. However, it seems that

---

CAUSE vs. HAVE, GO vs. BE, TCR (Terminal Coincidence Relation) vs. CCR (Central Coincidence Relation). Another proposal is that of Baker (2003), who assumes functional operators like Pred/BE, v/CAUSE, and Aspect/BECOME.

[80] According to Embick & Marantz (2008) –*ity* attaches to some specific roots like e.g. √ATROC and √CURIOUS and to adjectival-forming heads like –*able* and –*al*.

among the nº morphemes attaching outside v, for example, there may be different degrees of productivity and distribution. Let us consider the suffixes –*er* and –*ing*. They are both N-creating nodes merging with verbs but they do not compete with each other because they have different selectional (semantic) requirements, with the result that they have a different distribution. Then it seems that it is not that one has to ask for the productivity and distribution of N-forming terminal nodes, as Marantz claims, but for the exact restrictions that particular VIs are subject to. That is, one has to inquire about the domain of applicability of a particular suffix, in the sense of potential bases and affixes that it can be attached to. Another example of N-creating nodes are the suffixes -*ion* and –*ness*, which are inserted in different contexts and hence no competition is established between them: both suffixes are productive over the classes of bases they select (e.g. –*ness* selects the class of adjectives: $abstract_A > abstractness_N$ and –*ion* selects Latinate verbs: $conclude_V > conclusion_N$). They are not in competition (cf. Di Sciullo & Williams 1987, Williams 1994). Similarly, the suffixes –*miento* and –*ción* in Spanish are both nominalizing but they are used in different contexts: unlike –*ción*, -*miento* only attaches to verbal bases which contain the morpheme –*ec*-: *enriqu-ec-er* 'to enrich', *enriqueci-miento*, **enriqueci-ción* (Piera & Varela 1999: 4379, Varela 1999: 264). (See Williams 2007: 361-364, 364-367 for some criticisms against DM of their restricting competition to morphemes instead of VIs and of their associating contrasts of nominalization (e.g. the result being transitive or intransitive) with the presence of functional heads instead of the properties of the affixes themselves, e.g. –*ing*, -*ment* vs. –*ence*, Ø-suffix).

As was hinted at in the discussion of the previous question (i.e. competition between VIs), the denial of a lexicon or a kind of storehouse in DM is questionable. Let us consider Harley & Noyer's (1999) following quote (italics: ours):

(82)   "The content of a morpheme active in syntax consists of *syntactico-semantic features drawn from the set made available by Universal Grammar*".[81,82]

From (82), we infer that UG makes available morphosyntactic features and the underlying assumption is that all languages share the same set of features. However,

---

[81] An open question at the moment is to find out what the set of universal morphosyntactic features is.
[82] Similar statements can be found in other DM works (see e.g. Halle & Marantz 1993: 121).

some of them seem to be non-existent or have a zero-realization in some languages (e.g. gender). Could we hypothesize, based on this fact, that the pertinent (absent) feature is not present in a particular language at all? That is, if only some of the abstract morphemes which are taken from 'a universal feature inventory' are active in a particular language, it does not seem very economical that every time the speaker wants to use just some of them, they have to resort to the whole set of universal abstract morphemes.

A more economical option would be to have a kind of lexicon with those abstract morphemes specific to a language as well as the roots of such a language (i.e. features of a particular language would be selected from features of UG, much in the same way as in the MP, cf. Chomsky 1995 and subsequent work). This position would make more sense if roots are present with their phonological features from the beginning of the derivation (cf. subsection 1.4.1 where Late Insertion is discussed) and would certainly ease the process of language acquisition: children will not use any random root but only those roots specific to their language. If this position were a potential option, DM would then need a counterpart to a lexicon and their basic claim that there is no lexicon would not hold. A solution that DM could offer is to resort to the fact that language-specific features are not present in syntax but inserted at Spell-Out, which is in fact the option taken by DM followers.[83] On this view, all languages share the same set of syntactic features active in syntax and only some languages have specific featural requirements that are satisfied at Spell-Out. If this position is adopted, though, the presence of roots in syntax is still a problem. They are active in syntax but are specific to each language, i.e. they are not universal. The following quotes illustrate the point (italics: ours).

(83)   a. *"Roots are language-specific combinations of sound and meaning."*

Embick & Noyer (2007: 295)

---

[83] However, even this option seems to be questioned by DM followers themselves. Consider Harley's (2008b: 3) quote: "An Agr[eement] terminal node may be composed, *depending on the language*, of person, number, gender/class and case features" (italics: ours). See also the quote in (83b).

b. "(…) speakers of English memorize *Roots* such as √CAT or √SIT, as well as the fact that *abstract morphemes* such as [pl] and [past], which are drawn from a universal feature inventory, are *active in their language.*"

<div align="right">Embick & Noyer (2007: 296)</div>

If the reasoning up until now has some truth in it, then it seems that there must be a kind of storehouse/lexicon where the roots of a language can be placed. (Speakers cannot draw roots from any language). In some DM accounts, abstract morphemes active in a particular language would also be stored there.

To conclude, this subsection has presented some discussion around five claims made in DM: the alleged existence of some roots (i.e. the putative internal structure of some words), roots being acategorial, some flavours of category-creating terminal nodes, competition for insertion among VIs, and the denial of the lexicon. The discussion was not intended to settle the questions here but to open new questions for further research.

1.5 Conclusions

The present chapter started with an outline of Jackendoff's (1990, 1997, 2002) theory of grammar, known as the tripartite parallel architecture. Some arguments (e.g. syntax-phonology mismatches) were provided to support his view that syntax, phonology and semantics are three generative components, independent of each other, although connected by interface systems (cf. (9)). Although some of Jackendoff's arguments were not well-founded (or at least more evidence seemed to be required), as shown by works which propose a simple syntax-semantics interface (e.g. Baker 1985, 1988, 1997, Bouchard 1995, H&K 2002, Mateu 2002), there was still ample evidence for a non-uniform mapping between syntax and semantics. According to Jackendoff, each generative component contains word and phrasal structures. Authors like A&N (2004) have elaborated on such a model and have proposed that the word and phrasal subcomponents of phonology, syntax and semantics have their own vocabulary and principles of combination although some of them are also shared by the two subcomponents by their being inserted in big phonological, syntactic and semantic components (cf. (11)).

Section 1.2 was devoted to the syntactic component and its internal structure. It was shown that the claim that morphology can be accounted for by syntactic principles, and hence that there is no independent morphological component (cf. e.g. Sproat 1985, Baker 1985, 1988, Halle & Marantz 1993, Harley & Noyer 1999, Harley 2008b, a.o.) cannot hold. If morphology could be subsumed under syntax, there would be a simplification of the grammar, a desirable outcome. However, some evidence was provided for a morphological component, separate from the syntactic one, a view which has been present in the literature for some time (e.g. Di Sciullo & Williams 1987). In addition, some evidence was given for the generation of complex words (compounds included) in morphology and not in syntax. It was also shown that morphology and syntax, despite being separated into two subcomponents, share some vocabulary and principles, which is explained by being inserted into the same component (i.e. the syntactic one). We concluded that a model of grammar like that of Jackendoff (1990, 1997, 2002) or A&N (2003, 2004, 2007) in which morphology and syntax are separated (by each one heading its own submodule) but at the same time tied in some way (by being inserted in the same module) is flexible enough to capture the data.

Given the conclusions from section 1.2, the following section presented a morphological account of compounding. To be more precise, A&N's (2004) competition model was outlined and tested with some English and Romance (Catalan and Spanish) compounds. To put it briefly, A&N explain the existence of compounds by appealing to competition between syntax and morphology. On their view, a compound can exist in languages like English and Catalan/Spanish if it has no syntactic competitor, i.e. if it has a different semantics or a different merger of categories from its potential syntactic counterpart. It was shown that the morphosyntactic competition theory can explain most of the data examined here provided the semantic constraint assumed in the model (cf. (26)) was better characterized. Concerning the semantic constraint, we concluded that very subtle semantic distinctions between the morphological and syntactic structures must be taken into account for the competition model to work, delicate distinctions which were not included in the original proposal of A&N.

Despite the evidence provided for generating complex words in morphology, section 1.4 was devoted to presenting DM, a model of grammar according to which all word formation is syntactic (cf. Halle & Marantz 1993), especially due to the number of works proposing that morphology should be dealt with by syntactic principles. The

main theoretical assumptions of DM and some of its implications were discussed. Framed within this model, Harley's (2004, 2008b) analysis of compounds was explored. Some problems of her analysis were pointed out and some of her claims were questioned (e.g. the claim that internal arguments merge with the roots that select them before the roots are categorized). It was concluded that Harley's analysis should be thoroughly revised before it can account for the data satisfactorily; by contrast, an account along the lines proposed by A&N (2004) was seen as superior. For this reason, A&N's (2003, 2004, 2007) morphological approach to word formation has been chosen as the theoretical framework to explain the data which will be presented in the next chapters, where more will be said about the framework wherever that seems appropriate.

Chapter 2. Germanic and Romance compounding: the case of English and Catalan

This chapter starts (section 2.1) with some discussion about the notion of head in morphology: the validity of Williams' (1981a) Right-hand Head Rule (RHR) for English and Catalan morphology in general, and compounding more specifically, is established. Potential counterexamples to the RHR are explained away, and so are some arguments which have been raised in the literature to eliminate the notion of head in morphology (cf. Zwicky 1985, Bauer 1990 and Anderson 1992).

Section 2.2 is divided into two subsections: the first one (subsection 2.2.1) discusses the nature of the compounding elements in English and Catalan and the second one (subsection 2.2.2) contains a sketchy review of some compound classifications as they have been proposed in the literature to conclude that none of them is satisfactory enough. The only classification which looks promising is the one provided by Bisetto & Scalise (2005), which consists of two levels of analysis which are based on: (i) the grammatical relation between the compounding elements, and (ii) whether the compound is headed or headless (endocentric vs. exocentric compounds). These two levels of analysis give three big macro-types of compounds (subordinate, attributive, coordinate), each being subdivided into endocentric and exocentric.

Bisetto & Scalise's tripartite classification is adopted in section 2.3 to analyse the compound types available in English and Catalan and is provided with a further level of analysis, thus enriching the original classification. In the same section, after presenting the compounds following Bisetto & Scalise's scheme, our proposal follows. Coordinate compounds are claimed to be non-existent, and subordinate compounds and attributive compounds are argued to belong to the same underlying compounding type. Our proposal is based on English and Catalan data, but it is intended to apply generally. In short, the three macro-types, as proposed in Bisetto & Scalise (2005), are reduced to a single compounding type, an idea which is further developed in section 2.4, which also summarises the main results of the present chapter.

2.1 Some remarks on the notion of 'head'

This section is first devoted to establishing the notion of head in morphology (2.1.1). For such a purpose, Williams' (1981a) RHR is adopted and applied to English and

Catalan data, which is followed by some differences between morphological and syntactic heads. Subsection 2.1.2 discusses the proposals by Zwicky (1985), Bauer (1990) and Anderson (1992) against the notion of heads in morphology (or some part of it) and concludes that their arguments are not well-founded.

2.1.1 Heads in morphology

The notion of head, which plays an important role in syntax, can also be applied to the internal structure of words. Work on morphology has long established the existence of morphological heads (cf. Williams 1981a, 1981b Selkirk 1982, Scalise 1984, 1988, Di Sciullo & Williams 1987, Hoeksema 1988, 1992, a.o.) and most current work in morphology assumes their existence (e.g. Ackema 1999a, Bauer & Renouf 2001, Pérez Saldanya et al. 2004, Scalise 2008, a.o.).[84]

Of the criteria used for identifying syntactic heads, a number of authors agree that syntactic category is the relevant criterion, or at least one of the relevant criteria, for determining headedness in morphology (cf. Williams 1981a, Bauer 1990, Scalise & Guevara 2006). It is generally assumed that the head provides the construction of which it is a part with its lexical category through percolation, a mechanism which allows the syntactic category of the head to percolate up to the entire word, thus deriving its endocentricity (cf. Bauer 2003, Plag 2003, Booij 2005, Scalise & Guevara 2006; see also footnote 92).[85]

In the literature there is some dispute on how to identify the head. Based on the fact that morphological processes in English are typically right-headed, Williams (1981a: 248) proposes the Right-hand Head Rule (RHR) to identify the head in morphology. The RHR states that the head of a morphologically complex word is rightmost (see also Emonds 2006). Such a rule is meant to be applicable to both derivation and compounding.[86,87] A direct result of the RHR for derivation is that

---

[84] The pervasiveness of the notion of head in the morphological literature makes it impossible to list all relevant references. The ones listed here should be taken as a small sample of recent references which implicitly or explicitly make use of the notion of head.

[85] Note that the term 'head' refers here to the 'categorial head'. We are leaving aside the tripartite distinction among categorial, semantic and morphological heads made recently by Fábregas & Scalise (2008). For the time being, we will distinguish between a semantic head and a formal head. The formal head subsumes the categorial and morphological heads, the latter being responsible for features like gender and number.

[86] Authors like Booij (2005) argue that the source of the RHR has a historical explanation. Suffixes may have arisen from the second element of right-headed compounds, which in turn may have developed from right-headed phrases in languages whose syntax is right-headed (e.g. the suffix –dom in kingdom originates in the Old English word dom 'fate'). By contrast, prefixes usually emerge from a non-head

suffixes are predicted to be category-changing and prefixes category-neutral. The examples in (1) confirm such predictions for English ((1a) for suffixes, (1b) for prefixes). Concerning Catalan affixation, it is typically right-headed. Some examples are given in (2) ((2a) for suffixes, (2b) for prefixes).[88]

(1)   a. $mad_A+ness_N = madness_N$

          $character_N+ize_V = characterize_V$

          $beauty_N+ful_A = beautiful_A$

    b. $re+write_V = rewrite_V$

          $im+polite_A = impolite_A$

          $un+belief_N = unbelief_N$

(2)   a. $groc_A$ 'yellow' $+or_N = grogor_N$ 'yellowness/having the quality of yellow'

          $industrial_A$ 'industrial' $+itzar_V = industrialitzar_V$ 'industrialize'

          $brasil_N$ 'Brazil' $+er_A = brasiler_A$ 'Brazilian'

    b. $a+dormir_V$ 'to sleep' $= adormir_V$ 'to make somebody fall asleep'

          $anti+higiènic_A$ 'hygienic' $= antihigiènic_A$ 'antihygienic'

          $post+guerra_N$ 'war' $= postguerra_N$ 'postwar'

Leaving aside exocentric compounds for now, English compounding is also subject to Williams' (1981a) RHR. Accordingly, the rightmost formative within the compound will determine the category of the entire complex word (3). As for Catalan compounds, they are split into those which are right-headed (4a) and those which are left-headed (4b).[89]

---

position, be it the left constituent of a compound or a preverbal adverb (e.g. the English prefix *over-* in *overdo* comes from the independent lexical item *over*). A different account is given by Jackendoff (2007), who proposes that "the right-headedness of (English) compounds (…) really only relies on a language-specific correlation of linear order with semantic headedness, not on X-bar head-argument structure".

[87] The RHR does not seem to be universal. For example, Lieber (1981, 1983) notes that left-headed types predominate in Vietnamese and Thai, and Ceccagno & Basciano (2007) show that there is not a unique head position in Chinese compounds, which can be left-, right-, and double-headed. Hoeksema (1992) also agrees that a language may have more than one head position in the domain of morphology. The RHR must therefore be stated as part of the grammar of English. The RHR is then a parameter valid only for those languages with right-headed morphology or with a right-headed morphological subcomponent. See also footnote 89.

[88] The examples (1)-(2) here and (5)-(6) below are taken from Padrosa-Trias (2005b). Recall that a '+' sign is used for signalling the two elements of the affixed word (cf. footnote 4 in chapter 1).

[89] Scalise (1988: 243) also shows for Italian that there are right- and left-headed compounds, a division which, according to him, is based on the Latinate origin of the former and the synchronic native pattern of the latter. Clements (1992) also observes a division of headedness in Spanish compounds, for which he

(3)   [black$_A$+board$_N$]$_N$

   [jet$_N$+black$_A$]$_A$

   [computer$_N$+generate$_V$]$_V$

(4)   a. [cama$_N$+trencar$_V$]$_V$ (leg+break) 'to break the leg(s)'

   [mal$_{Adv}$+interpretar$_V$]$_V$ (badly+interpret) 'misinterpret'

   [cama$_N$+curt$_A$]$_A$ (leg+short) 'short-legged'

   b. [camió$_N$+cisterna$_N$]$_N$ (lorry+tank) 'tanker lorry'

Not only is the head of a word necessary for formal reasons (e.g. category determination) but also for semantic reasons: the compound is a hyponym of the head (cf. the 'IS A' relation, Allen 1978: 105). This semantic test becomes especially relevant for identifying the head in those cases in which the two constituents of the compound are of the same category like *camió cisterna* in (4b). The two words forming the compound being nouns, one could argue that it is also the rightmost noun which determines the category of the compound. However, by the hyponymy criterion, the compound as a whole is a hyponym of *camió*, not of *cisterna*. (See subsection 2.2.2 for other pieces of evidence to identify the head in a compound).

Although the RHR seems to apply quite consistently in the pertinent morphological subcomponents, there are some data which are in conflict with it and question the claim that the head in morphological constructions is on the right. In compounding, the RHR is difficult to hold for exocentric, i.e. headless, compounds like *pickpocket*. Although one could argue that the noun *pocket* is the (categorial) head that determines the nounhood of the compound (see e.g. Fábregas & Scalise 2008 for Romance compounds), it cannot be the (semantic) head since the compound does not denote a kind of *pocket*. The two elements that make up the compound are attributed to (predicated of) an entity (a person) which lies outside the compound. One could, nonetheless, assume that there is a zero-suffix embodying the missing entity. If such an approach is correct, the RHR could still be maintained for such exocentric compounds: the zero-suffix would be responsible both for the category and the semantics of the compound. Note that the zero-suffix proposal is not necessary for other traditionally

---

proposes a Lefthand Head Rule and a Righthand Head Rule for left- and right-headed compounds respectively.

considered exocentric compounds like *faintheart* if metonymic processes are allowed to operate. The noun *heart* can give the nominal category to the compound (categorial head) and is the semantic head of which *faint* is predicated (i.e. the heart is faint, which in turn is predicated of a person: a person who has a faint heart). (The validity of the RHR for Catalan compounding will be seen in sections 2.3.2 and 2.4).

In the domain of derivation, there are at least two types of data which challenge the RHR, each of which will be dealt with in turn. The first challenge is posed by category-changing prefixes. For example, Williams (1981a) observes that the English prefix *en-* systematically converts nouns and adjectives into verbs, thus displaying the behaviour of a head:

(5) $rage_N$ > $[en+rage]_V$
        $noble_A$ > $[en+noble]_V$

A similar scenario exists in Catalan. The prefix *en-* also seems to convert nouns and adjectives into verbs in a productive way.

(6) $caixa_N$ 'box' > $[en+caixa+ar^{90}]_V$ 'to put (something) in boxes'
        car_A 'expensive' > $[en+car+ir]_V$ 'to raise the price (of something)'

Other putative category-changing prefixes in English are the following ones: *a-* as in $[a+[sleep]_V]_A$, and $[a+[kin]_N]_A$, *be-* as in $[be+[friend]_N]_V$, and $[be+[calm]_A]_V)$, and *de-* as in $[de+[bug]_N]_V$ (cf. e.g. Marchand 1969, Siegel 1979, Williams 1981a, Selkirk 1982, Bauer 1990, Anderson 1992, Carstairs-McCarthy 2002). Other allegedly verbalizing prefixes in Catalan include *a-* as in $[a+[genoll]_N+ar]_V$ (A+knee+IS) 'to kneel (down)', *re-* as in $[re+[fred]_A+ar]_V$ (RE+cool+IS) 'to cool (down)', and *des- (es-)* as in $[des+[coratge_N]+ar]_V$ (DES+courage+IS) 'to discourage', and $[es+[teranyina]_N+ar]_V$ (ES+cobweb+IS) 'sweep (spider's webs)' (cf. Cabré & Rigau 1986, Cabré 1988, 1994).

In front of these counterexamples to the RHR, one is faced with different alternatives to explain them. One option is to assign the attribute of a head to the prefix and have left-headed prefixed words (cf. e.g. Siegel 1979, Williams 1981a, Fabb 1984,

---

[90] The final suffix in the examples in (6), i.e. *–ar* and *–ir*, and the final suffix in all the Catalan prefixed words in the following paragraph, i.e. *–ar*, is an inflectional suffix (IS) which indicates that the verb belongs to a particular conjugation (they belong to the first conjugation except for *encarir*, which belongs to the third one).

Bauer 1990, Gavarró 1990a, Hoeksema 1992, Lieber & Baayen 1993). If correct, such a view destroys the RHR's prediction that syntactic category identifies the morphological head in a systematic fashion. A second alternative to deal with the counterexamples to the RHR is not to treat them as exceptions, which is the view defended by authors such as Scalise (1984, 1988), Neeleman & Schipper (1992), Gràcia (1995), Stiebels (1998), and Padrosa-Trias (2005a, b, 2006, in press, a); they argue - for a number of different Romance and Germanic languages - that prior to prefixation there is a conversion process of adjectives and nouns to verbs, by means of a zero-affix.[91] Some evidence for this conversion-analysis comes from the argument structure of verbs, assuming that the Θ-grid of a complex word is derived from the thematic information of its morphemes via Θ-role percolation (see the original works for details). The view that category-changing prefixation is just apparent can only be maintained if the assumption of Θ-role percolation with respect to the RHR is clarified, which leads us to the second challenge to the RHR.

The basic idea of Θ-percolation is that the thematic information of a complex word is derived from the different elements that form the word, irrespective of whether they are prefixes or suffixes. This view of Θ-percolation is in conflict with the RHR, which states that only the head is able to transfer its features. Other problematic data come from prefixed verbs in which the prefix changes properties of the base verb to which it attaches, a possibility which should be disallowed by the RHR. For example, the prefixes in *disabuse* and *dispossess* change the syntactic subcategorization frame of the base verb: the prefixed verbs have an argument introduced by the preposition *of*, which is not present in the base verb (for other similar counterexamples to the RHR, see Bauer 1990: 23-29). Such data show that the strict RHR (Williams 1981a) has to be abandoned, in favour of the Rel(ativized) RHR (Di Sciullo & Williams 1987: 25-28), according to which the head for a specific feature is the rightmost element that contains the feature in question. Such a revised definition of the RHR can account for the percolation of Θ-roles and of syntactic subcategorization frames which come from

---

[91] Zwanenburg (1992a) also reaches the same conclusion but without postulating a zero-suffix. He adopts Walinska de Hackbeil's (1985) analysis in which words prefixed by the alleged category-changing prefixes in English (e.g. *be-*, *en-*, *de-* and *a-*) form a PP but differs from hers in that the PP does not change into a V by means of a zero-suffix, but by the conversion rule V → PP.

different relativized heads within the word (i.e. rightmost heads with respect to the feature they contribute to the complex word).[92]

Concerning the information which has been claimed to percolate from relativized heads to the entire complex word, there have been several proposals. Not only the syntactic category, Θ-roles and syntactic subcategorization frames have been assumed to percolate, but also other features such as the [+/-animate] feature (Scalise 1984), the [+/-Latinate] feature (Lieber 1981, Williams 1981a), gender in languages like German which have grammatical gender (Lieber 1981), features marking tense, aspect, person and number (Williams 1981a), and theta-grids and case features (Fabb 1984).[93] We view percolation as the mechanism responsible for the transmission of the aforementioned features (e.g. gender) from the (relativized) head to the top of the morphological tree. Such a view does not pose any problem to the Rel. RHR (Di Sciullo & Williams 1987) and will prove useful for the classification of compounds (sections 2.2.2 and 2.3). Although the term *percolation* may remind one of the 1980s, its current use is still valid and is comparable to other mechanisms which have been proposed recently. For example, Neeleman & van de Koot (2002a) propose a syntactic mechanism which allows both the transmission of argument structure and features like syntactic category, by means of upward copying of functions and features introduced by terminal nodes. Such syntactic process can be easily adapted to morphology and can be seen as a modern version of a mechanism similar to percolation.

---

[92] For other early statements of feature percolation, see e.g. Selkirk (1982), Fabb (1984), and Lieber (1983, 1989). They all have the same effect as Di Sciullo & Williams's (1987) proposal but with different terminology.

[93] What kind of information is able to percolate is a matter of debate. For example, Lieber (1989) argues that argument structure should not be passed upwards via percolation but via inheritance. Inheritance refers to the relationship between the argument structure of a derived word and its input elements. A complex word inherits an argument from the base when the argument may be represented as an argument of the derived word either syntactically (sometimes referred to as external or syntactic inheritance) or internally to the complex word (sometimes called internal or morphological inheritance). To see the effects of inheritance, let us consider a concrete example. In the derived adjective *washable*, the internal argument of the base verb *wash* is inherited and is realized as the external argument of the complex word. Compare:

(i)   a. I wash the red towels
         b. The red towels are washable

Whereas it is generally agreed that inheritance accounts for the shared thematic structure between (ia) and (ib) (cf. e.g. Booij 1988, Levin & Rappaport 1988, Picallo 1991, Spencer 1991, Neeleman & Schipper 1992, Gràcia 1992, 1995, Gràcia et al. 2000), there are also some claims that point to another direction. In this respect, Hoekstra & van der Putten (1988) and Bordelois (1993), among others, prefer to talk of a shared semantic structure, not of strict inheritance, between the two lexical items.

In short, all potential counterexamples to the RHR in the languages under discussion can be explained by the relativized notion of head, as formulated by Di Sciullo & Williams (1987). The Rel. RHR will be adopted in what follows. The notion of morphological head which has emerged from the previous discussion is different from the notion of head that exists in syntax (see below for further differences), which means that heads in syntax and morphology are simply not identical: they are defined by different criteria (although some of them may be shared by the two distinct heads). This view follows from the fact that syntax and morphology are two independent submodules with their own principles within a bigger syntactic module (see the model of grammar depicted in (11) in chapter 1). Recall from chapter 1 (section 1.2) that there are other pieces of evidence that support the claim that heads in morphology are different from heads in syntax, which in turn is evidence for the separation of morphology from syntax. For example, headedness in morphology and syntax is not regulated by the same principles. Whereas morphological heads in English are systematically on the right, syntactic heads are typically on the left. The situation in Catalan is more complex but the morphology and the syntax of Catalan are also subject to different principles (see subsection 1.2.1 in chapter 1 for details) (cf. Plag 2003).

2.1.2 Against heads in morphology

Despite the widely accepted claim that the notion of morphological head is a valid one, some authors question it either in the whole domain of morphology, such as Zwicky (1985) and Bauer (1990), or in some submodule of it, such as Anderson (1992).[94] Such questioning has its main source in the attempt to apply to a complex word the exact same set of criteria that is usually used to determine the head of a construction in syntax, which gives conflicting results as to what should be the morphological head (cf. Zwicky 1985, Hudson 1987; the latter cited in Bauer 1990).

Now Zwicky's (1985), Bauer's (1990), and Anderson's (1992) arguments that seem to invalidate the notion of head in morphology will be sketched first, some of which (e.g. alleged category-changing prefixes) have already been presented and will not be discussed again. Then, it will be shown that their arguments against the existence

---

[94] It is outside the scope of this thesis to discuss a number of other theories which also deny the existence of heads and explain regularities in word formation by other means (e.g. via a neural network; cf. Rumelhart & McClelland 1986, see Pinker 1999 for a good summary of their position). To illustrate the point here it will suffice to sketch three authors' proposals, namely Zwicky's (1985), Bauer's (1990) and Anderson's (1992), which deny the existence of heads in some fashion.

of morphological heads are not well-founded. Finally, it will be concluded that heads exist both in syntax and morphology but that they are subject to different conditions. The two heads need to be distinguished, as has been argued in the previous subsection.

2.1.2.1 Zwicky (1985)

Zwicky (1985) concludes that heads have a very limited role in morphology. In his view, the notion of head can only apply to endocentric compounds. He reaches this conclusion on the following grounds. According to him, the morphological head must fulfil two roles: (1) be the morphosyntactic locus (i.e. bear the inflectional markers), and (2) determine the category of the word.

As for the first requirement, he argues that there is no need to refer to the head of the word to explain the presence of inflectional suffixes in derivation and compounding, but rather to the margins of a word, thus explaining plural forms such as *maple leaves* and *baby teeth*. So far there is no difference between Zwicky's view and Di Sciullo and Williams' (1987) Rel. RHR (or Williams' 1981 RHR for this purpose). Some data which can tell them apart is the contrast found between the plural formation of *sabertooth* and *baby tooth*. In Zwicky's view, the plural of *tooth* should be the same for both compounds, i.e. **saberteeth* and *baby teeth*, contrary to reality, given that in his view inflectional affixation occurs at the margins and is indifferent to the internal structure of the word. By contrast, the Rel. RHR predicts that plural marking will be realized on the rightmost element of the word (specified for this feature). *Baby tooth* being endocentric, *tooth* constitutes the righthand element of the compound. As a result, the compound will change to *baby teeth* when pluralized. *Sabertooth* could, in principle, receive the same treatment: *tooth* could be analysed as the head which is given the attribute of being like a *saber*, and via metonymic processes *sabertooth* refers to a type of animal. However, we believe that speakers perceive *sabertooth* as a simplex word, and consequently, there is regular plural formation: *sabertooths*. In conclusion, the expression of inflection in morphology requires the internal structure of a word and the notion of head, and not that of margin of a word.

Concerning the second requirement the morphological head must satisfy, i.e. category determination, Zwicky concludes that there is no consistent semantic notion which can be applied to both derivation and compounding: the semantic functor (i.e. the suffix) determines category in derivation whereas categorization is usually determined by the semantic argument (i.e. the rightmost element) in compounding. What these two

semantic notions have in common, though, is the position in the word, namely the rightmost position within the complex word in each case, thus giving further support to Williams' RHR. Contrary to Zwicky, heads in morphology are then necessary. They are the locus of morphosyntactic features and they determine the category of the complex word when it is not exocentric.

2.1.2.2 Bauer (1990)

Bauer (1990) reaches similar conclusions to Zwicky. Bauer applies a set of criteria, that a prototypical head fulfils in syntax, to morphology (see Bauer 1990: 2-3 for the list of criteria, the source of which is Zwicky 1985 and Hudson 1987). As already noted, the criteria give diverging results as to what constitutes a head in morphology. To illustrate the point, the distribution criterion will be considered as it applies to suffixation in English. It will be seen that even a single criterion does not converge on what should be taken as the head. In Bauer's (1990: 2) terms, "The head of a phrase is the distributional equivalent of the whole phrase (this is Bloomfield's criterion once more)". When such a syntactic definition is extended to morphology, to class-maintaining suffixes, the distribution test does not provide clear results. Whereas some suffixed words have the same distribution as the base (e.g. compare *greenish* and *duckling* with *green* and *duck* respectively), others do not (e.g. *kingdom* does not have the same distribution of the base, nor its suffix (Bauer 1990: 8)).

Given that the distribution test in syntax does not work in the same way as it works for suffixation and that other criteria used for determining headedness in syntax fail to identify a consistent head in morphology, Bauer (1990: 30) concludes that "*heads have no place in morphology*. Certainly, if they have a role to play, this role needs to be defined much more carefully than has been the case up until now" (italics ours). However, as observed earlier in this section, the notion of head in morphology is necessary. It must then be the case that syntactic and morphological heads cannot be defined by the same set of criteria. Whereas some tests which determine headedness in syntax are also applicable to morphology (e.g. category determination), others may not work for a subcomponent of morphology, like suffixation (e.g. the distribution test). Another difference between syntactic and morphological heads is that heads are identified hierarchically in syntax (X-bar theory in generative grammar, e.g. GB theory) and positionally (Rel. RHR) in morphology.

2.1.2.3 Anderson (1992)

Contrasting with most current analyses which assume that words are internally structured and that the notion of relativized head explains the properties of the word as a whole, Anderson (1992) accepts both facts for compounds only (with some amendments, though) and denies them for derivation. Anderson explains that the notion of head is necessary in some compounds to explain the irregular inflection of plural forms, such as *scrubwoman/scrubwomen*, and of past tense forms in cases like *outdo/outdid* (cf. *sabertooth/sabertooths* and *baby tooth/baby teeth*). Whereas he admits that compounds may have internal heads, he sees no need for them to have a fixed position within the complex word, thus denying rules like Di Sciullo and Williams' (1987) Rel. RHR.

To account for headed derived words, Anderson is forced to enlarge the class of words which have internal structure, which he terms "composites". Composites are words with internal structure and include compounds, e.g. *scrubwoman* and *outdo*, what he calls combining forms like *Sino-Japanese* and *erythromycin*, and prefix-stem combinations of the type *receive* and *conceive*. Anderson differentiates compounds, which are the product of syntactic rules, from both combining forms and prefix-stem combinations, which are stored in the lexicon with their internal structure visible, their parts not occurring independently. Anderson claims that such accessible structure allows them to have the desired allomorph (e.g. *ceive ~ sep*, as in *receive ~ reception*) and be used as an analogy for the creation of other forms. If one accepts Anderson's classification into composites and derived words, it remains to be seen whether his claim that neither internal constituent structure nor heads exist in derivation can be maintained. As (7) shows, Anderson himself admits the existence of certain structure internal to affixed words.

(7)   "On the view that words have *no* non-phonological [i.e. morphological] structure which is accessible to other rules, all principles for the placement of affixes ought to be purely phonological. (…) but there are some instances in which the morphology of the form seems to be relevant to affix placement, such as the *person-marking prefixes of Georgian Verbs* that are attached directly to the Verb stem in a way that ignores an aspectual prefix if one is present. There is *no phonological definition of this position*: the person markers go at the very beginning of the non-prefixed forms, regardless of their syllabic or other

phonological structure, and are only 'infixed' if the form contains an aspectual prefix (…) *we must evidently admit a limited presence of structure-building operations in Word Formation.*"

<div align="right">Anderson (1992: 304; italics ours)</div>

Similarly, despite his denial of heads in affixed words, at some points he compares the head of a derived word with the entire word (see the passage below), which to our understanding suggests that he implicitly acknowledges the existence of heads, thus contradicting his initial claim:

(8)   "(…) it seems that affixation never needs to identify any non-phonological aspect of a word's structure except *to specify its domain as being the head of the word as opposed to the entire word*. It thus appears that an organization of some words into a *head and a non-head periphery* is all the structure that is warranted in the output of Word Formation Rules. This is still not ideal (since it would obviously be preferable to prohibit structure-building altogether), but at least it does not imply that every Word Formation Rule that applies leaves its structure behind for later rules (potentially) to access."

<div align="right">Anderson (1992: 305; italics ours)</div>

By denying to affixed words internal structure and heads, Anderson explains that the properties of words usually come from the last suffix by means of Word Formation Rules (WFRs) (compare the classic 'Adjacency Condition' of Siegel 1978 or 'Atom Condition' of Williams 1981a). The order in which they apply gives the linear sequence of suffixes in a word that contains more than one suffix. The rightmost suffix corresponds to the last WFR whose effects are visible since no other WFR has applied afterwards. His conclusion that affixed words have no internal heads does not follow. What follows, instead, is the fact that the properties attributed to internal heads are not visible to word-external operations.

As it stands now, Anderson´s theory faces a number of shortcomings. Baker (1993) and especially Carstairs-McCarthy (1993) provide a thorough discussion of the strengths and weaknesses of Anderson's a-morphous morphology.[95] Only some of

---

[95] See Halle & Marantz (1993) for other criticisms against Anderson (1992).

Carstairs-McCarthy's arguments in defense of the existence of internal structure and of heads in morphology will be provided below. (The reader is referred to the original works for details).

First, Anderson's attempt to explain *able*-suffixed adjectives via truncation of verbs with an *–ate* suffix (e.g. *demonstrable < demonstrate*) by appealing only to phonology fails. Given that stress placement is not identical in all varieties of standard English, Anderson's claim that *–ate* truncation only applies when it has no primary stress does not hold. The phonological account, for example, leaves unexplained why *truncable* and *translable* do not exist in American English, given that the suffix *–ate* does not bear primary stress (*trúncate, tránslate*). Similarly but in the opposite direction, Anderson's account cannot explain the existence of words like *mutable* (compare *truncable* and *translable*) in British English, if primary stress is placed on *–ate* in such a disyllabic verb in British English. The conclusion is that the phonological approach does not make the correct predictions: it both overgenerates and undergenerates, which makes such an approach untenable and calls for another explanation. It seems that the morphological structure of words is necessary: *-ate* truncation is only possible when *–ate* has suffixal status in the language (see Carstairs-McCarthy 1993: 214-215 for details).

In the same vein as the previous argument, Anderson attempts to reduce to phonology the absence of the past participle prefix *ge-* on German verbs with 'inseparable prefixes', which are unstressed. Anderson claims that *ge-* is not sensitive to whether the verb to which it attaches to form the past participle is prefixed or unprefixed, but rather to whether it is initially stressed or not. The absence of *ge-* follows from the absence of initial stress, so Anderson claims for German: both prefixed verbs *bespréchen* 'discuss' and nonprefixed verbs *riskíeren* 'risk' lack *ge-*. Such a phonological account does not extend to Dutch, a language from which the same behaviour would be expected because it possesses the same ingredients as German. Dutch can prefix *ge-* to initially-unstressed unprefixed verbs (e.g. Dutch past participles *geriskéerd* 'risked', *geexaminéerd* 'examined'). The presence and absence of *ge-* in Dutch past participles seems to be tied to the absence and presence of an inseparable prefix on the verb respectively, i.e. to the internal morphological make-up of the word and not to the stress properties of the verb, thus denying Anderson's a-morphous morphology, according to which morphemes are illusory artifacts, not real elements in word structure.

Furthermore, if the irregular plural and past tense inflections of *scrubwoman/scrubwomen* and *outdo/outdid* are accounted for by appealing to the notion of head, so must derived words like *undo, pre-sell* and *rewrite*, whose past tense forms are *undid, pre-sold* and *rewrote*, and not **undoed, *preselled* and **rewrited*, as would be expected from Anderson's account, according to which derivation lacks internal structure and hence the notion of head is irrelevant. Carstairs-McCarthy notes that in this case analogy cannot save such prefixed words, given that analogy in Anderson's view only applies to composites whose constituents are made up of bound forms. The irregular past tense of such prefixed forms can then be explained iff their internal structure is visible and the head can be identified. These facts destroy the initial distinction Anderson draws between compounding and other complex words, the former allegedly having internal structure while the latter lacking it.

Carstairs-McCarthy observes that the same reasoning can be extended to cover the German verbs of the sort *besprechen* 'discuss' and *erfinden* 'invent, discover', discussed earlier. They have irregular inflection; for example, for *erfinden*: *erfand* and *erfunden*. Again, on Anderson's analysis, regular inflectional forms are expected if there is no internal structure, and hence no head. To explain the irregular forms of such verbs, Anderson makes use of "paradigm preservation" (using Carstairs-McCarthy's terminology, p. 223). There is no apparent reason, other than that of affirming internal constituency, for paradigm preservation not to apply to Anderson's composites such as *receive* [siv ~ sept] and *scrubwoman* [wYm↔n ~ wImIn]. In other words, resorting to different treatments for essentially the same data seems a high price to pay to keep his original distinction between composites - which Anderson assumes to have internal structure and heads - on the one hand, and derivation - to which Anderson attributes no internal structure and no heads - on the other.

In short, the division between composites, which Anderson claims have internal structure and where heads have a role to play, and the rest of complex words, with no internal structure and hence no heads, has proved rather artificial. The conclusion is that heads and internal structure are necessary in all complex words.

In conclusion, this section has confirmed the existence of heads in morphology, which, although similar to syntactic heads in some respects, need to be distinguished from them (contra Zwicky 1985, Bauer 1990, and Anderson 1992). We have seen that syntactic and morphological heads are not identical, i.e. they are subject to different

conditions, which explains the failure of applying to morphology the criteria which are typically used to identify the syntactic head. As will be seen in the next section, the notion of head will play a crucial role in the classifications of compounds.

2.2 What are compounds and how to classify them

Definitions of compounds as well as classifications for them have been objects of debate in the literature for a long time, and yet no satisfactory definition and classification seem to be available. The main goal of subsection 2.2.1 is to identify the categories of the two compounding elements in English and Catalan compounds. The main body of the discussion will be centered on English compounding, which, if not stated otherwise, will also apply to Catalan compounding. For example, the terminological chaos surrounding the definitions of compounds is present both in English and Catalan, but, in order to illustrate such chaos, definitions of compounding as they apply to English compounds are selected. Subsection 2.2.2 provides a brief review of several classifications of compounds that have been proposed in the literature. Again, for ease of exposition, the exemplification of the different proposals is based on classifications intended to explain English compounding. However, it seems reasonable to extend them to Catalan compounding since similar compounding classifications are also available for Catalan.

2.2.1 The raw material of compounds

Definitions of what a compound is abound in the literature. Thus, spotting a compound should be a relatively easy task, but this is not always the case, mainly due to the different terminology used in the definitions. Some authors view compounds as the result of putting together two roots (9), two stems (10), two lexemes (11), two words (12) and two bases (13), the last one including a combination of some of the previous terms. We agree with Bauer (2001: 695) when he states that "the category 'compound' is very poorly defined", despite being a common phenomenon cross-linguistically. (See Olsen 2000b: 897-898 for a summary of how the concept of compounding has evolved since its inception; see also Lieber & Štekauer 2009 for some recent discussion on how

to define a compound). Below are some quotations which illustrate the terminological chaos (italics in (9-13): ours).[96]

(9)   a. "(…) compounds, that is words formed by combining *roots*, (…)"

Carstairs-McCarthy (2002: 59)

b. "Compounding occurs when two independently meaningful *roots* are directly combined to form a new, complex word, usually a noun or adjective."

Harley (2006: 99)

(10)   a. "(…) compounding [is concerned] with the formation of new lexemes from two (or more) potential *stems*."

Bauer (1983: 33)

b. "If two *stems* are sisters (i.e. they form a compound), (…)"

Lieber (1983: 253)

(11)   a. "Compounding is a process by which a **compound lexeme** is derived from two or more simpler *lexemes*."

Matthews (1991: 82)

b. "The formation of a new lexeme by adjoining two or more *lexemes* is called **compounding** or **composition**. Nearly all languages have compounds and, in many languages, compounds are the main type of new *lexeme*."

Bauer (2003: 40)

c. "(…) compounding processes in which novel lexemes are formed from the combination of two simpler *lexemes* (…)."

Spencer (2003b: 329)

---

[96] Definitions parallel to the ones given in (9) to (13) for English are also found in the literature devoted to compounding in Catalan. For example, compounds as the union of roots is the view defended by Mascaró (1986: 22); compounds as the combination of stems is the position held by Cabré & Rigau (1986: 134), Duarte & Alsina (1986: 9) and Gràcia (2002: 781); and compounds as the putting together of two lexemes is Cabré's (1994: 83) view. Vague definitions making reference to words and bases are also available. For instance, Gavarró (1990b: 113) views compounding as the "concatenation of words" and Pérez Saldanya et al. (2004: 247) understand compounding as a word formation process which combines two bases.

d. "In many languages, compounding (also called **composition**) is the most frequently used way of making new *lexemes*. Its defining property is that it consists of the combination of *lexemes* into larger words."

<div align="right">Booij (2005: 75)</div>

(12) a. "A 'compound word' is usually understood to be the result of the (fixed) combination of two *free forms,*[97] *or words* that have an otherwise independent existence, as in *frostbite*, *tape-measure*, *grass-green*."

<div align="right">Adams (1973: 30)</div>

b. "A word whose parts may themselves be *words* in other contexts is traditionally called a **compound**."

<div align="right">Matthews (1991: 14-15)</div>

c. "A compound is a word which consists of *two or more words*."

<div align="right">Fabb (1998: 66)</div>

(13) a. "Bases may be free elements, able to occur on their own, or they may be bound forms with no independent existence, as in words like *dental*, *holism*, *amorphous*, whose bases have meanings like those of English words –'tooth', 'whole', 'form'. Bound bases will be referred to as stems. *Bases*, both *words* and *stems*, may combine to form compounds: *credit card*, *oviraptor*, *pesticide*."

<div align="right">Adams (2001: 2)</div>

b. "(…) a **compound word** contains at least two *bases* that are both *words*, or at any rate, *root* morphemes."

"(…) a prototypical compound is a word made up of at least *two bases* which can occur elsewhere as independent *words*, for instance, the compound *greenhouse* contains the bases *green* and *house* which can occur as words in

---

[97] See Bloomfield (1933: 227), who also gives a vague definition of compounds in terms of free forms: "Compound words have two (or more) free forms among their immediate constituents".

their own right (e.g., in the noun phrase *the green house*, i.e., the house that is green).”

<div align="right">Katamba & Stonham (2006: 55, 304)</div>

c. “(…) a compound is a word that consists of two elements, the first of which is a root, a word or a phrase, the second of which is either a root or a word.”

<div align="right">Plag (2003: 135)[98]</div>

Finding out the exact nature of the compound elements is necessary for a complete classification of compounding in a language. Two elements put together may be a compound or not depending on the definition adopted. For example, the Danish form *cigar+mager* ‘cigar maker’ and the English form *war+monger* will not be considered compounds under the definition that compounds are made up of two (or more) lexemes (cf. 11), because whereas in both cases the first element can exist on its own (i.e. it is a lexeme), the second element cannot. However, if compounds are understood as a form containing two potential stems (cf. 10), then both the Danish and English complex words fall under the definition of being a compound, because *mager* and *monger* can take inflectional markers, such as the plural and genitive markers, as in *cigarmagere* (plural) and *warmonger’s* (genitive) (see Bauer 1983: 38-39 for more examples and discussion of this point). The identification of the units used to form compounds is also relevant for testing some hypotheses put forth in the literature, such as the hypothesis by Snyder (2001) (cf. chapter 3) about the alleged correlation between the presence of a certain construction and the availability of a compound type in a language.

After acknowledging the relevance of identifying the status of the elements forming compounds, each definition, or rather each group of definitions, will be submitted to close scrutiny, to reach (hopefully) a unique and uniform definition of what a compound consists of. We will start with the definitions given in (12), according to which compounding is a matter of putting two words together. Note that the term ‘word’ is itself ambiguous. Matthews (1974) distinguishes three senses: (i) the lexeme, the fundamental unit listed in the lexicon, which can be exemplified by the lexeme SMILE which can have forms like *smiles*, *smiling* and *smiled*; (ii) the word-form, also known as the orthographic word; and (iii) the grammatical word, according to which the

---

[98] Plag (2003: 10) first defines compounding as the combination of two bases to form new words. Later on (cf. 13c) he explains what he understands for bases.

word-form *smiled* is the past tense or past participle of the lexeme SMILE.[99] Katamba & Stonham (2006) make the same three-way distinction, but define lexemes as the vocabulary items which are listed in the dictionary, i.e. what Di Sciullo & Williams (1987) have called listemes. These listed units may include both morphological and syntactic objects, objects stored in the lexicon due to some idiosyncratic property (e.g. the syntactic unit *kick the bucket* has unpredictable semantics in its idiomatic interpretation and hence will be listed in the lexicon). Harley (2006) shares the same view as Katamba & Stonham: lexemes and listemes have a related meaning, but they are not interchangeable terms. Harley defines listemes as "the units that encode a sound-meaning connection –they are the things that are listed in the mind of the speaker (…)" (p. 111). In short, lexemes and listemes should not be fused together, but should rather be kept separate. Lexemes, and not listemes, are the entities relevant to the present discussion: they are the elements linked to one sense of the term 'word' mentioned by Matthews (1974) and it is precisely lexemes that are being referred to by the definitions given in (12), the defining unit also present in (11). Before discussing whether lexemes are the real objects behind compounds, we will turn to another umbrella term, that of 'base', used in (13).

The term 'base' is a cover term for two, or three, units: words and stems according to Adams (13a), words and roots according to Katamba & Stonham (13b) and Plag (13c), the latter adding phrases as another plausible unit in the nonhead position. Once we realize that the term 'word' is a vague word which really stands for 'lexeme', we are left with roots, stems and lexemes (leaving aside phrases for the moment), as the units for compounding, the units which are also claimed to be relevant for compounding in (9), (10) and (11) respectively. The first apparent labyrinth of definitions of what a compound is made of is getting more manageable. That compounds in English and Catalan can be formed by lexemes is clear from examples like *(to) hand-wash*, *grass-green*, *high school* (English) and *un contra+atac* (a counter+attack) 'a counterattack' and *faldilla pantaló* (skirt trousers) 'skort' (Catalan). What remains to be seen is the difference between root and stem compounds and to establish their existence if that is the case.

Giegerich (1999) claims that stems have no place in present-day English morphology. Within a base-driven stratification model, he identifies two morphological

---

[99] Lexemes are conventionally written in block capitals and word-forms in italics.

categories for English: the root and the word (the lexeme in our terms). For him, roots can be free or bound, and are not specified for syntactic category; words are free forms and are specified for category. Giegerich defines stems as bound bases which carry syntactic category and denies their existence for current English. He argues that in earlier stages of English such a category had a *raison d'être*, given that the language had a rich inflectional system, but since the decline of such a system, the category 'stem' has been lost or become indistinguishable from that of word. Thus, English is left with a root- and word-based system. A three-way system is found in German, where, in addition to roots and words, stems have a role to play. The bases to which the German suffixes –*bar* and –*ung* attach are stems (e.g. *trinkbar* 'drinkable' and *Schöpfung* 'creation'): they are verbs but are not ready to enter syntax because they need the appropriate inflection (see Giegerich 1999: chapter 3 for details). Similarly, the need for a stem level in German, unlike English, is seen in stem-composition: the adjective *red* in *roter Wein* (red wine) is inflected because it modifies the noun in syntax, but is uninflected when it forms part of a compound: *Rotwein*. A stem level is also needed in Catalan. There are also cases of verbal bases which cannot appear in syntax unless some affixation takes place. That is the case of the bases to which the suffix –*able* attaches: *imaginable* 'imaginable', *recomanable* 'advisable' and *llegible* 'readable'. We also find cases of stem compounding in Catalan: verbs like *mal+gastar* (badly+spend-INF) 'to waste money', *contra+dir* (counter+say-INF) 'to contradict' and *espanta-sogres* (scare-PRES/STEM[100](…).3SG+mothers-in-law) 'party blower' cannot occur in syntax unless they take the appropriate verbal inflection. As these compounds illustrate, stems in Catalan compounding can occur in first position (*espanta-sogres*) and second position (*mal+gastar*, *contra+dir*).

Accepting Giegerich's (1999) claim that stem-based compounds are non-existent in current English and having confirmed their existence in Catalan, we now have to (dis)confirm the existence of root compounds. In this respect, there are at least two different views. On the one hand, the authors (Carstairs-McCarthy 2002, Harley 2006) who describe compounding as the union of two roots make a distinction between free and bound roots to account for ordinary compounds, which are based on free roots of the language, and those formations like *biology* and *television* (compounds in their view), which are based on Latin and Greek and are specially used for scientific and

---

[100] As will be seen when discussing the nominal VN compounds in Catalan, there is no agreement on the nature of the first element.

technical vocabulary. If we accept such classical-based formations as compounds on a par with ordinary compounds, then the term 'root' seems to be a handy one to describe compounding in English. That is, the process of compounding in English could uniformly be described by means of a single term, i.e. 'root'. What this view of compounding fails to realize, though, is that free roots are in fact lexemes, words ready to enter syntax, which destroys the apparent uniformity of this view of compounding. On the other hand, those authors (Matthews 1991, Bauer 2003, Spencer 2003b, Booij 2005) who describe English compounds as the result of putting two lexemes together, obligatorily need another term to account for classical-based formations, since the two units forming such formations are not lexemes, i.e. they are not ready to enter syntax, hence creating non-uniformity in the process of English compounding. This conclusion, though, is not necessarily the only one: classical-based formations behave differently from ordinary compounds and one might argue that they do not belong to the category of compounds. Then, English compounding could also be described uniformly by means of the term *lexeme*. Observe that adopting either position has no repercussions on Catalan compounding, which cannot be defined uniformly by means of a unique category due to the existence of stem compounds. By looking at some data from the classical languages, we will consider next whether uniformity is possible in the case of English compounding, and if so which position is more satisfactory. It will be concluded that uniformity, despite being a desirable property, is not possible.

If forms like *geography*, *telephone*, *hydrology*, *theology* and *bureaucrat* are compounds, they are of a rather peculiar sort. They are found in English and some other European languages but the elements are usually Greek and Latin in origin. This is why they are known as neo-classical compounds. Some authors prefer the term 'combining form' (CF), and distinguish between initial combining forms (ICFs) like *astro-* and *electro-*, and final combining forms (FCFs), like *-crat* and *-phobe*. As Katamba & Stonham (2006) observe, the status of such classical-based forms depends on the speaker's knowledge of the classical languages. A form like *hydrology* might have no internal structure for a speaker with no knowledge of Greek, but might be analysed as the union of two bases (*hydro* 'water' and *logy* 'science or study') for a person with some knowledge of Greek[101]. The meaning of the two elements can alternatively be reconstructed from the meaning of the complex lexeme as a whole, and from its

---

[101] See also Pérez Saldanya et al. (2004: 248-249) for a similar view: learned compounding, as they call it, is only available to those speakers who are part of a specialized and technical field.

comparison to other forms sharing the same CFs and identifying the recurrent meaning in the two complex forms.

At first sight, one could argue that such CFs resemble affixes in that both can be added to lexemes (cf. Wheeler 1977: 246-247). Both a prefix and an ICF can be added to the lexeme *electric* (14) and both a suffix and a FCF can be added to the lexeme *music* as well (15).

(14)   a. an+electric
        b. photo+electric

(15)   a. music+al
        b. music+ology

<div align="right">Bauer (1983: 213)</div>

Despite this similarity in behaviour, there are other properties that differentiate CFs from affixes. The fact that CFs can occur in more than one position in the word (compare **theo**logy and poly**theism**) speaks against treating them as affixes. If they were affixes, it would be odd to have one which can be both a prefix and a suffix. Surprising also would be the possibility of combining two such forms, like in *theocracy* (god+rule), because it is well-known that prefixes and suffixes need a base to create a well-formed lexeme: **de-ist*, **mis-let*. All this suggests that these classical–based formations should not be treated as affixes and could be regarded as compounds.[102]

Although CFs can be combined with lexemes (cf. 14, 15), like ordinary compounds, they typically combine with bound roots (e.g. *glaciology*, *vibraphone*), unlike ordinary compounds. Untypical of compounds is also the linking vowel between the two elements of neo-classical compounds. As has been noted by several authors (Scalise 1984: 75-76, 99, Oniga 1992: 110-111, Adams 2001: 118), the vowel changes depending on whether it is a Greek or Latin compound, being usually *–o* in the former (e.g. *heterodoxia* 'other, different opinion') and *–i* in the latter (e.g. *aurifer* 'gold-bearing'). Plag (2003: 157-158) observes that there is no linking vowel if the ICF ends in vowel, e.g. *sui-* as in *suicide*, or the FCF starts with vowel, e.g. *–itis* as in

---

[102] See Buenafuentes (2007: 357-360) and Varela (2005: 74) for other points according to which CFs and affixes behave differently, e.g. their semantic contribution to the complex word is different.

laryngitis.[103] In addition, at the start of discussing neo-classical compounds, we mentioned that most of them are academic and technical terms, frequently used in specialised fields, but not so often used by the average speaker who finds, for example *nephrotomy* (*nephr- nephros*, 'kidney' and *tom-, tomos*, 'cutting') much harder to understand than *kidney-cutting*, the more transparent English counterpart of the Greek-based complex word (example from Adams 1973: 131).[104] The differences between ordinary compounds and neo-classical compounds discussed so far suggest that they are two different processes of word formation in English, their formations being regulated by different principles.

In short, we have seen that there are a number of distinct properties that distinguish neo-classical compounds from ordinary compounds but that at the same time they are closer to compounding than to affixation. We have to conclude then that alongside the major, lexeme-based system, English has a root-based one, the latter being reserved for neo-classical compounds. It seems that neither position sketched at the beginning of this discussion (i.e. compounding being all root-based or all lexeme-based) can be maintained, although the second one seems preferable. The first view fails to notice that free roots are in fact lexemes and that English prefers free roots to bound ones in everyday speech. The second view fails to account for neo-classical based formations if English compounding is all lexeme-based, which can be considered less problematic if such formations are taken as a special subgroup of compounds. The same observations can be applied to Catalan.[105] Given that neo-classical compounds are subject to different word formation principles from ordinary compounds, they will not be taken into account in the classification of compounds in English and Catalan (section 2.3).

After revising the definitions given in (9-13) and reaching the conclusion that lexemes and roots are needed to account for compounding in English, and that lexemes, roots and stems are required to explain compounding in Catalan, the last point which

---

[103] Some neo-classical compounds vary in form. Sometimes the variation is meaningful, as is the case in *strati-* 'stratum' and *strato-* 'stratus', which refer to rock and cloud respectively but in many other cases the change in form does not seem to bring along a change in meaning, as is the case in *pulsimeter* and *pulsometer*, *toxidermic* and *toxoprotein*, *dosimeter* and *dosemeter*, and *spermaduct* and *spermicide*. Still in other cases a shorter and longer forms coexist: *dermo-* and *dermato-* (cf. Adams 1973: 130-131, Adams 2001: 119).

[104] For the special phonological behaviour of neo-classical compounds, which further distinguishes them from ordinary compounds, see e.g. Plag (2003: 156-157), Carstairs-McCarthy (2002: 66).

[105] On this point, see Cabré & Rigau (1986: 154-155), Mascaró (1986: 77-84), Cabré (1994: 82-87) and Gràcia (2002: 824-825).

needs to be discussed in this subsection is whether phrases can also form parts of compounds (cf. 13c: Plag 2003).

For a long time it was believed that phrases could not occur in a compound. The 'No Phrase Constraint', as formulated by Botha (1981), prohibits the occurrence of phrases inside words, compounds included (16). Such a constraint may be a language-specific requirement, though: at first sight (and according to traditional analyses of Catalan compounding) it seems to be valid for Catalan in both head and non-head positions (which is the position which will be presented below but see our treatment of some Catalan compound types, namely traditionally considered coordinate compounds, in subsection 2.3.2 for a different view) but it is not valid for English in the non-head position.

(16)   "Morphologically complex words cannot be formed (by WFRs) on the basis of syntactic phrases."

Botha (1981: 18)

There are some data that apparently contradict the general assumption that phrasal compounding is not available in Catalan. That is, there are cases in which a word and a phrase are combined in a compound, as in *[[menjador]$_N$-[sala d'estar]$_{NP}$]$_N$* 'dining.room+sitting.room'. On closer examination, we see that the phrase *sala d'estar* is in fact a lexicalized phrase and that no freely generated phrase can be inserted in this compound. That is, phrasal compounding in Catalan (probably Romance in general; see Gaeta 2006 for Italian) seems to be severely restricted and consequently the 'No Phrase Constraint' seems to be valid. Also, in Romance some authors refer to lexicalized phrases like *sala d'estar* as 'synaptic/syntagmatic' compounds because of their phrasal nature. On our view, they are not compounds at all but simply lexicalized syntactic phrases and hence are not considered in the study of compounding in Catalan. To avoid confusion, 'phrasal compounding' will be used to refer to what is found in Germanic languages, i.e. the insertion of a phrase in a compound (see below) and 'synaptic/syntagmatic compounds' will refer to what is found in Romance languages, i.e. lexicalized syntactic phrases (more will be said on this point at the end of the subsection of nominal compounds in Catalan, subsection 2.3.2.1).

The English counterexamples to the constraint expressed in (16) are done away with by arguing that they are lexicalized or frozen phrases (similar to the position we

have just considered for Catalan). This is the view defended, for example, by Carstairs-McCarthy (2002), to explain the presence of phrases in compounds and derived words, like (17):

(17)   [fresh air] fanatic
        [open door] policy
        [French histori] an
        [nuclear physic]ist
        [sexually transmitted disease] clinic

                                                Carstairs-McCarthy (2002: 81-82)

Carstairs-McCarthy summarizes his position as follows: "(…) lexically listed phrases (i.e. idioms) or institutionalised ones (i.e. clichés) can appear in some contexts where unlisted phrases cannot." (p. 82).[106] A similar view is held by Bresnan & Mchombo (1995: 194), for whom phrasal compounds are also lexicalized and taken as quotations, which can include phrases from foreign languages, among other material (see also Wiese 1996). Such a view can explain the following examples:

(18)   a [mea culpa] look
        a certain [je ne sais quoi] quality
        the [ich bin ein Berliner] speech

However, not all phrases inside compounds can be considered lexicalized material, like idioms, clichés, and loan phrases. Bauer (1983: 164) and Lieber (1992: 11) provide

---

[106] This is also Allen's (1978) position. She explains the phrases in (i) by claiming that they are "mini-idioms which must be listed in the permanent lexicon with their non-compositional meanings just like sentential and verbs-phrase idioms (p. 238)." On this account, the ungrammaticality of the examples in (ii) is expected, given that they are not idioms, but transparent phrases semantically.

    (i)   [black and blue]ness
            [at-home]ish
            [out-of-doors]y
    (ii)   *[intelligent and attractive]ness
            *[at school]ish
            *[open woods]y

Also note that authors like Spencer (2000: 318) consider that "This type of construction is more tolerable in English when the phrase can be perceived as lexicalized (in some rather unclear sense), (…)".

examples like (19) and (20) respectively, which clearly show that non-conventionalized phrases can also appear as left-hand members of compounds:[107]

(19)   a [don't-tell-me-what-to-do]$_{CP}$[108] look

        an   [oh-what-a-wicked-world-this-is-and-how-I-wish-I-could-do-something-to-make-it-better-and-nobler]$_{CP}$ expression

        a [pain-in-stomach]$_{NP}$ gesture

(20)   a [pipe and slipper]$_{NP}$ husband

        [off the rack]$_{PP}$ dress

        a [slept all day]$_{VP}$ look

        a [pleasant to read]$_{AP}$ book[109]

        [over the fence]$_{PP}$ gossip

Not all phrases in the non-head position are possible, though. Whereas *a [French history] teacher* is fine, **a [the French history] teacher* is not, which shows that definite NPs with a determiner are not a possibility (cf. Booij 2005: 79).[110]

---

[107] Two real cases of phrasal compounding we overheard in 2008 are given below:

(i)   (…) the regularish [official research group] coffee break
(ii)   This is going to turn into one of those really cool [small world I know your friends] thing!

[108] We will not go into the details of the layered CP since it is not the point of the present discussion.
[109] Sato (2007) provides similar examples, some of which follow:

(i)   a nice [easy-to-drive] car
        [hard-to-imagine] behaviours
        [difficult-to-solve] equations

Despite not being the main concern here, just note that phrases can be used not only as the non-head of a compound, but also as the basis to form a verb and as the base to which a derivational suffix can attach.

(ii)   He [I-don't-care]-ed his way out of his room
        Patrick [I-don't-take-that-kind-of-crap-from-anyone]-ed his brother
                                                        Carnie (2000: 91, cited in Sato 2007)
(iii)   I feel particularly [sit-around-and-do-nothing]-ish today.
        This is definitely a blower-upper, not a [leave-it-where-it-is]-er
                                                        Bauer (1983: 70, 71)

[110] Meibauer (2007: 237) challenges the commonly held assumption that the non-head cannot be a DP. He observes that "where the non-head is a fixed expression, DPs are possible", which he exemplifies with the following German data:

(i)   die [der-schöne-Rheingau]-Laberei (the the-beautiful-Rheingau talk)
        die [Ein-Kerl-wie-ich]-Visagen (the a-guy-like-me faces)   Wiese (1996: 191)

Unlike Romance languages where phrasal compounding is not possible (or severely limited), in Germanic languages phrasal compounding is in fact quite productive. Meibauer (2007) argues for German that phrasal compounds are not marginal at all. Taking into account the relevance of hapaxes as a sign of productivity, he concludes that phrasal compounds must be productive because most of them are hapaxes and just a few are lexicalized. He also shows that the phrase in the non-head position is not restricted to a single category, but a range of phrasal categories are allowed (e.g. NP, PP, VP, CP), like in English. Some of his examples are given in (21).[111]

(21)   der [Zehn-Tage]NP-Urlaub (the ten-days holidays)

        die [Vor-Premieren]PP-Fahrt (the before-première trip)

        die [Länger-leben]VP-Diät (the longer-live diet)

A&N (2004: 124) reach the same conclusion for Dutch:

(22)   [waarom leven wij?]CP probleem (why live we problem)

        [blijf van mijn lijf]VP huis (stay-away from my body home)

        [ijs met slagroom]NP fobie (ice-cream with whipped-cream phobia)

        [bozer dan boos]AP blik (angrier-than-angry look)[112]

Note that despite having placed a phrasal category (e.g. NP, AP, VP, CP) next to each phrase in the non-head position of the compound, they are all treated as simplex words once they are integrated into the compound (see A&N 2004 for a matching mechanism between the phrase and the terminal node in which it is inserted to form the compound; see also footnote 20 in chapter 1). On a superficial level then, (despite appearances) they can be regarded, for example, as NN compounds. While acknowledging the existence of English phrasal compounds of the type illustrated in (19-20), they will not be included in the classification of compounds in English for ease of exposition (cf. subsection 2.3.1). Given that they can be assimilated to non-phrasal compounds, nothing will be lost by not considering them.

---

[111] See Botha's (1981: 73-76) examples of compounds whose non-heads also show a wide range of phrasal categories.
[112] Thanks are due to Peter Ackema for providing me with this example.

Next, we have to consider whether phrasal embedding is also a possibility in the head position of the compound, a possibility disconfirmed by the examples in (23) with the interpretation of a compound:

(23)   *white [big board]

        *telephone [long conversation]

        *filing [modern cabinet]

        *knife [great sharpener]

Whereas this possibility does not exist in English, other languages like Dutch can make use of it, as A&N (2004: 124) show:

(24)   namaak [mobiele telefoon] (imitation mobile phone)

        rot [luie stoel] (rotten comfy chair)

        dat kanker [Juinense accent] (that cancer Juinen accent) 'that bloody J. accent'

To summarize this subsection, the raw material of English compounds is (i) a root, a lexeme or a phrase in the non-head position, and (ii) a root or a lexeme in the head position, reserving roots for neo-classical compounds. As for Catalan compounds, the compounding elements can be a root, a stem or a lexeme both in the head and non-head position, roots also being present in the case of neo-classical compounds only. (In subsection 2.3.2, we will see that some phrases can also occur in the non-head position of some Catalan compounds). Having clarified the terminology surrounding compounds, the terms 'word' and 'base' will be used in what follows when the real category behind these cover term is irrelevant. When necessary, the more specific term will be chosen.

2.2.2 Which is *the* classification of compounds?

It has often been noted that basing one's study of compounding on traditional works on the topic (e.g. Marchand 1969, Adams 1973, Bauer 1983) may not cover all patterns of compounding in the language. Traditional works tend to study well-established compounds of the language, which are usually tied to the idiosyncrasies that the passage

of time[113] usually brings along (e.g. semantic drift), and they cannot take into account new emerging patterns of compounding. Recently the relevance of neologisms has been stressed in the field of word-formation (e.g. Baayen 1992, Bauer & Renouf 1996, Plag 2003, Meibauer 2007), with the subsequent growing number of authors basing their studies of compounding on large corpora, which can (dis)confirm old existing patterns and identify new emerging ones (cf. e.g. Berg 1998, Bauer & Renouf 2001, Bauer 2004, Bisetto & Scalise 2005, Scalise & Guevara 2006, Ceccagno & Basciano 2007).

To take into account the crucial role played by neologisms and corpora, the different types of compounding available in English and Catalan (cf. subsection 2.3.1 and subsection 2.3.2 respectively) are based on a comparison of different sources: classical works (e.g. Allen 1978, Selkirk 1982 for English and Moll 1952, 1975, Fabra 1956, Badia 1962, Mascaró 1986 for Catalan), recent textbooks, handbooks and papers on the topic (e.g. Spencer 1991, Bauer 2003, Lieber 2003, Plag 2003, Booij 2005 for English and Cabré 1994, Gràcia & Fullana 1999, 2000, Adelman 2002, Gràcia 2002 for Catalan) and corpus-based studies (e.g. Bauer & Renouf 2001, Ceccagno & Basciano 2007, for English).[114] Such a comparison will allow us to represent the current patterns of compounding, disregarding those which were once present but no longer are and considering those which were once absent but present nowadays. Such a procedure will also prevent us from developing a biased classification, of which there is more than one example in the literature, mainly due to the fact that a lot of work on compounding has focused on primary/root compounds on the one hand, and on secondary/verbal(-nexus)/synthetic compounds on the other, thus neglecting many other compound types. Such a result is mainly due to the extensive number of works on English compounding.

Before presenting some compounding classifications available in the literature, we feel it is necessary to clarify some terminology. Nowadays many authors use the terms 'synthetic' and 'verbal' compounds interchangeably to refer to compounds in

---

[113] Compounding will be studied from a synchronic point of view only. In other words, the development of a free lexeme, which once was used to form compounds, into a suffix will not be considered (e.g. that is the case of the suffix –*hood* in the derived words *childhood* and *parenthood*, the suffix no longer being associated to a free lexeme). By contrast, the passage of a suffix to a lexeme (e.g. *ism, ology, emic, etic,* cf. Bauer 1983) used to form compounds will be taken into account. However, examples of the latter type were not found when carrying out the survey of compounds in English. Note that authors like Borer (1989, 2008) define compounds in Hebrew as "groupings of two nouns which result in idiosyncratic, non-productive meaning (…) (1989: 48)", a view which we reject.
[114] To our knowledge, corpus-based studies devoted to Catalan compounding similar to the ones available for English are missing. Such a lack has been counteracted by using native speakers' judgments.

which the second element has a verbal base and the first element is interpreted as an argument of the verb (e.g. Bauer 2003, Lieber 2003, Plag 2003, Katamba & Stonham 2006, Selkirk 1982). This position, though, is not held by Botha (1984; cited in Bauer & Renouf 2001), for example, who distinguishes the two terms. Also, other scholars like Roeper & Siegel (1978) only include under the rubric of 'verbal' compounds those whose deverbal noun finishes in -er, -ing, and –ed. Still other authors like Bisetto & Scalise (2005) regard 'synthetic' and 'secondary' compounds as synonymous. We will treat 'secondary/verbal/synthetic' compounds as synonymous terms and include all (de)verbal nouns, irrespective of the nominalizing suffix, e.g. *task assignment*, *crime prevention* and *body massage*, as members of such compounds. For a summary of the different positions adopted in the area of synthetic compounding see e.g. Bauer (2001: 701-702). Note that 'argument' is sometimes understood in broad terms: "an element bearing a thematic relation such as Agent, Theme, Goal, Source, Instrument, etc., to the head" (Selkirk 1982: 23).

Concerning 'primary/root' compounds, they are understood as those compounds whose head is not (de)verbal or whose non-head is not an argument of the (de)verbal head. Compounds like *climbing equipment*, *mass production* and *fitness campaigner* would be examples of such compounds (cf. Carstairs-McCarthy 2002: 63). By looking at the constituents of the examples just given, the term 'root' looks quite inadequate to characterize such compounds, and for this reason the term 'primary' will be used in this thesis and for uniformity the term 'secondary' will be chosen from 'secondary/verbal/synthetic' compounds (but see footnote 129).

After this terminological clarification, we will review several classifications that have been proposed in the literature in order to see which one is the best one to accommodate the existing compounding patterns of the languages under study.[115] Classifying compounds, especially nominal compounds, in terms of their meanings has proved rather difficult. Below, some proposals (Lees 1960, Hatcher 1960, Levi 1978, Downing 1977) - most of which are based on a set of semantic categories - will be

---

[115] Recall that the classifications we will review here are intended to explain English compounding. Similar conclusions, nevertheless, can be drawn from classifications of compounds in Catalan. For a review of some of them, see Gavarró (1990b: chapter 2). To illustrate the point, Lees' (1960) analysis presented here, for example, is comparable to Wheeler's (1977) treatment of Catalan compounds: they both derive the surface form of compounds from a syntactic deep structure.

sketched to show that none of them provides a satisfactory enough account.[116] After reviewing such semantic-oriented approaches to compounding briefly, other proposals taking into account factors other than semantics will be considered (Bauer 1983, 2003, Booij 2005, Plag 2003, Carstairs-McCarthy 2002, Bisetto & Scalise 2005).[117]

Within the standard theory of Transformational Grammar (Chomsky 1957, 1965) the lexicon contains simple words only, and compounds (and derived words) are explained by transformational rules. On this background Lees (1960) proposes that, by means of a number of transformations, nominal compounds like *man-servant* are derived from an underlying sentence (roughly: 'the servant is a man').[118] The grammatical (syntactic) relation between the constituents of nominal compounds is the same relation holding in the underlying structure from which the compound is derived (e.g. subject-predicate in the case of *man-servant*). Lees explains the fact that some compounds can have more than one interpretation by having different deep structures from which the different interpretations arise. For example, the compound *elephant bed* can be interpreted differently depending on the deep structure: e.g. (i) The bed has the shape of an elephant, (ii) The bed is for the elephant, (iii) The bed has the picture of an elephant drawn on it, (iv) It is a bed to be thrown at elephants when they are wild, among the countless possibilities (whose only requisite is that there must be some connection between *bed* and *elephant*). As can be deduced from the previous paraphrases, Lees' proposal has some problems, some of which have been amply discussed in the literature since Chomsky's (1970) seminal work "Remarks on Nominalization" (cf. e.g. Gleitman & Gleitman 1970, Matthews 1974, 1991, Bauer 1983, 2003, Scalise 1984). We will limit ourselves to point out just a few flaws of such a transformational account, which will make it clear that such an approach is untenable.

First, a large number of verbs that enter into the structural description of the transformation can be deleted, which gives too much power to the transformations. For instance, *power* is deleted from 'wind powers the mill' in *windmill*, *grind* from the 'mill grinds flour' in *flourmill*. As can be observed from the examples just given, there is no consistent transformation for NN compounds.

---

[116] For recent semantic-based approaches to NN compounding, see Benczes's (2006) study within the theoretical framework of Cognitive Grammar and Jackendoff's (2007) study in his Parallel Architecture model.

[117] Although Lees derives compounds by a series of syntactic transformations, these are based on the semantic paraphrases given to compounds, which explains why Lees has been included within the semantic approaches to compounding.

[118] See Lees (1960: chapter 4) for the different transformational rules he proposes to derive nominal compounds from deep sentences.

A second point is that a possible paraphrase for *wind mill* is, as noted in the previous paragraph, that 'wind powers the mill', which can be taken as the underlying structure from which *wind mill* originates. The deletion of the verb *power* would explain the surface form of the compound. However, *windmill* can also be paraphrased as 'the wind activates the mill', 'the wind makes the mill function' and 'the mill is activated by the wind' (cf. Scalise 1984: 12). In each case, a different verb is deleted, and a different transformation is needed. Despite there being different paraphrases, at least they are not incompatible, but note that the compound *wind mill*, when not being used in its conventional meaning, can also be used to describe 'a mill in which one always feels a lot of wind', 'a mill in which there are lots of pictures depicting windy weather', 'a mill from which one can see the wind blow'. That is, a single surface compound may have semantically incompatible interpretations. Another example is *Eskimo dog*, noted by Gleitman & Gleitman (1970), for which they give the following interpretations: 'dog used by Eskimos', 'dog that looks like an Eskimo', and 'dog that lives in igloos'. The possibility of having countless possible readings means that an endless number of different underlying syntactic structures are needed if the compound is to be derived from such a structure (with implausible operations to delete all the syntactic material).

Third, Lees' unrestricted view of transformations and deletions also leads to the problem of recoverability. In Chomsky's (1965: 138) terms, "only recoverable deletions are permitted". That is, it is not possible to delete lexical material which cannot be recovered afterwards. Such a requirement on transformations is not observed by Lees' proposal: an indefinite number of verbs can be deleted, and given the number of paraphrases a compound may have, a set of verbs, and not just one, can be deleted from any given compound, with the consequence that deleted material is often not recoverable.

Another criticism to the transformational account is that the sentential origin of compounds cannot explain some idiosyncrasies found in compounding. The existence of *boyfriend* and *girlfriend*, but not *adultfriend*, *manfriend* and *childfriend*, of parallel form, requires the pertinent transformational rules to be blocked in these particular cases, not expected under the assumption that transformational rules are regular and systematic. It is also difficult to imagine which deep sentences would underlie compounds like *striptease* and *cuptie* (cf. Matthews 1974: 188-194).

On the basis of this quick review of the transformational account of compounding, which rests on a now obsolete theoretical framework, one cannot

conclude that Lees' approach is not the one to follow in the classification of compounds nowadays, but it has helped us see some aspects which current word-formation theories should avoid. For example, irrecoverable material should not be present in a theory of word-formation, and yet we find some recent syntactic proposals in which this very same problem is still present, like Harley's (2002, 2004) (see subsection 1.4.2.3 of chapter 1 and especially Padrosa-Trias 2007a for a review of its problems). It is on the basis of these more recent syntactic analyses of compounding that the point made in chapter 1 is reinforced, i.e. that compounding (word-formation in general) cannot be explained by syntax and that morphology cannot be dispensed with. In short, syntax and morphology are two independent systems, each with a different domain of application (cf. the model depicted in (11) of chapter 1).

Semantic-based approaches which use a set of semantic categories for explaining the possible compound patterns and their meanings do not agree on the number of categories which should be distinguished. The final number of distinctions really depends on how fine-grained one's analysis is. On the one hand, some studies try to come up with as many distinctions as possible in order to cover all imaginable semantic relations between the two elements of the compounds (e.g. Brekle 1970, in Plag 2003). On the other hand, other studies try to come up with broad and more abstract categories so that only a few of them will be necessary to accommodate all compounds (e.g. Hatcher 1960, Levi 1978, Downing 1977). An immediate problem with the maximalist approach is that, due to the large number of semantic distinctions, some of them overlap, and an immediate problem with the minimalist approach is that due to the small number of distinctions, the semantic categories are very general, with the need for further sub-divisions in some cases.[119] Otherwise sometimes one does not know where to place some compound types. For example, Hatcher only has four main categories for nominal compounds, which are based on the relation between the two members of the compound (e.g. A being the first element and B the second one: 'A is contained in B' as in *gold ring*, 'B is contained in A' as in *broomstick*, 'A is the source' as in *cane sugar*, and 'A is the destination' as in *New York express*), but then she recognizes the need for further subdivisions for all compounds to fit in a specific category. She introduces seven subcategories, which are based on the reference of each

---

[119] A minimalist approach to compounding would probably treat *book-keeper* and *office-worker* in the same way, as 'Noun plus Agentive', whereas a maximalist treatment would make finer distinctions: *book-keeper* would be 'Object plus Agentive', and *office-worker* would be a member of 'Locative plus Agentive' (cf. Matthews 1991).

element (e.g. 'person', 'animal', 'place', 'time') and which can occupy both positions in the compound resulting in 49 subdivisions. Even then, some compounds are placed in a category with some difficulty, i.e. not always may it be clear where to class a particular compound (see Hatcher 1960: 369, 373 for difficult examples).

Deriving nominal compounds from underlying relative clauses (comparable to Lees's 1960 approach to compounding), Levi (1978: 76-77) identifies nine possible relationships, which she claims are recoverable: 'cause' (*drug deaths*), 'have' (*apple cake*), 'make' (*silkworm*), 'use' (*pressure cooker*), 'be' (*soldier ant*), 'in' (*morning prayers*), 'for' (*headache pills*), 'from' (*sea breeze*) and 'about' (*sex scandal*). As Levi puts it (p. 76): "This set is made up of 9 predicates: CAUSE, HAVE, MAKE, USE, BE, IN, FOR, FROM and ABOUT. These predicates, and only these predicates, may be deleted in the process of transforming an underlying relative clause construction into the typically ambiguous surface configuration of the CN [Complex Nominal]". If some compound does not fit into any of the previous relationships, then the relation between the two elements of the compound must be overtly expressed to be comprehensible, as in *moss-covered rocks*, but as Adams (2001) notes this is not the right conclusion because compounds like *moss house* are indeed allowable and fully understood. Adams also notes that some compounds involve a relationship other than those listed by Levi and are nevertheless fully comprehensible: *a speed table* is understood as *an anti-speed table* (cf. Adams 2001: 82-89 for these and other examples and for a review of different semantic approaches to NN compound interpretation).

Downing (1977) represents a departure from earlier approaches to nominal compounding in the sense that she rejects the idea that compounds are derived from an underlying syntactic structure (cf. Lees 1960, Levi 1978) and emphasizes the fact that studies on compounding should not be based on familiar, oft-cited compounds, like most previous studies, but on novel, non-lexicalized compounds which are absent from idiosyncratic features that time brings along (e.g. different degrees of lexicalization).[120] Downing (1977) recognises that context plays a fundamental role to identify the exact relationship holding in a compound out of the infinite relations which are in principle possible (cf. Borer 2009). She lists twelve semantic classes which according to her correspond to the most common relationships underlying the constituents of a nominal

---

[120] The study of lexicalized compounds may not be a good indicator of existing compound types as the patterns on which the lexicalized forms are based may have become obsolete and hence speakers may not be able to create new forms on such pattern.
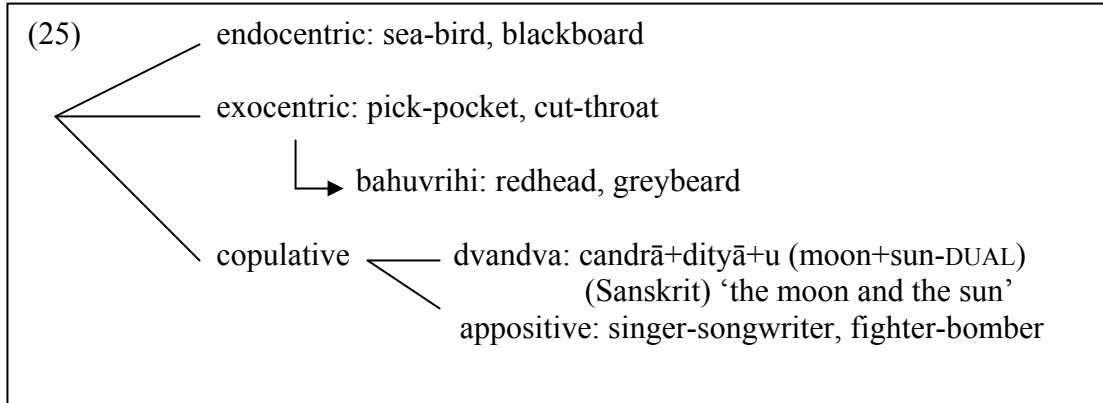
compound (p. 828): whole-part (*duck foot*), half-half (*giraffe-cow*), part-whole (*pendulum clock*), composition (*stone furniture*), comparison (*pumpkin bus*), time (*summer dust*), place (*Eastern Oregon meal*), source (*vulture shit*), product (*honey glands*), user (*flea wheelbarrow*), purpose (*hedge hatchet*) and occupation (*coffee man*).

Downing's classification, like the previous semantic-based classifications, gives rather arbitrary results. A recurrent problem in all semantic-oriented approaches to compounding is that they all give a different number of semantic categories without justifying it. The question of why there are no more or less categories than the ones each analysis proposes is not answered in any approach. As already noted before, some compounds may be difficult to place, resulting in heterogeneous groupings sometimes (e.g. *exclusion zone*, *World Cup*, *water-power* vs. *horse-power*) and other compounds may be placed in more than one category. For example, in Levi`s classification, *rabbit warren* could be classified as 'have' (the rabbit has a warren), 'in' (the rabbit is in the warren) and 'for' (the warren is for the rabbit). They are not exclusive categories. In addition, experimental work like Devereux & Costello's (2007) has shown that establishing the relation between the constituent members of the compounds is not enough to correctly interpret novel NN compounds. They suggest that relation-based approaches to compounding should be combined with concept-based approaches, with the result that the intrinsic properties of the compounding concepts are also taken into account in the interpretation.[121]

Due to the rather heterogeneous and arbitrary semantic categories found in semantic-based treatments of compounding, semantics cannot be used as the sole basis for classifying compounds. As will be seen next, the semantic relationship between the two members has also been taken into account in other compound classifications (giving the categories of dvandva and appositive compounds, see (25)), which have, in addition, considered other aspects, such as the presence or absence of a head (hence the categories endocentric vs. exocentric compounds). This is the approach taken by Bauer (2003)[122] and Booij (2005), who come up with the following classificatory scheme for compounding:

---

[121] Downing (1977: 831) already hints at such idea when she observes that speakers classify compounds differently depending on whether the head of the compound refers to "naturally existing entities (plants, animals, and natural objects)" or "synthetic objects", the former being typically classified in terms of their inherent features and the latter in terms of the purpose for which they are created (see also Pustejovsky 1995).

[122] Note that Bauer (1983) treats his main four types of compounds in his classification as semantic types which can interact with the syntactic category of the whole compound and the syntactic category of the

```
(25)        endocentric: sea-bird, blackboard

            exocentric: pick-pocket, cut-throat

                    ⌐→ bahuvrihi: redhead, greybeard

            copulative ── dvandva: candrā+dityā+u (moon+sun-DUAL)
                                    (Sanskrit) 'the moon and the sun'
                          appositive: singer-songwriter, fighter-bomber
```

Before discussing the adequacy of such a classification, let us see some differences between Booij's and Bauer's own classifications. First, Bauer (1983, 2003) fuses exocentric with bahuvrihi compounds whereas Booij (2005) observes that bahuvrihi compounds are sometimes treated as a subgroup of exocentric compounds (cf. 25). A reason to separate them is that bahuvrihi compounds (also known as possessive compounds, cf. Plag 2003 below), and not exocentric compounds, involve two types of predication: a predication of quality which is established between the two elements of the compound (e.g. the beard is grey in *greybeard*) and a predication of attribution which is predicated of an entity outside the compound (e.g. of a person whose beard is grey in the case of *greybeard*) (cf. Benveniste 1974: 155-159). Note also that the formation of compounds of the *redhead* type and the *pickpocket* type is regulated by different principles. Many compounds of the *redhead* type have body parts in the second position, which are modified by the first elements: *fathead, boldface, greybeard, pale-face, redbreast* and *redskin*, most of which have a corresponding adjectival form in *−ed*, as in *redskinned*. The possibility of having extended forms is not available to the *pickpocket* type of compounds (Marchand 1969, Spencer 1991, Adams 2001). (More will be said in subsection 2.3.1, in the classification of English compounds.)

Second, Bauer (1983, 2003) also treats copulative and dvandva[123] compounds as one and the same category, and the compounds included under 'appositive' are not given a name in Bauer (1983, 2003) but in Bauer (2001, 2008) they are called appositional and are treated as another type similar to copulative/dvandva compounds but not exactly the same. The term 'copulative' has been used in different senses by

---

elements forming the compound, an approach more similar to Plag (2003). Although Bauer (2003) and Booij (2005) may assume this kind of interaction, they do not make it explicit.
[123] 'Dvandva' is the Sanskrit name for a coordinate compound.

different authors, which explains why Olsen (2000b, 2001) argues that copulative compounds are productive in English whereas according to Bauer (2001), they are not common. Such opposing views arise from the fact that Olsen refers to the compounds listed under 'appositive' in (25) and Bauer to the compounds listed under 'dvandva'. Examples similar to the Sanskrit dvandva given in (25) are only found in English in borrowed place-names (e.g. *Alsace-Lorraine*) and corporate names (e.g. *Time Warner*), according to Bauer (2001),[124] although their status as dvandvas is also controversial. For example, Olsen (2001) argues that real dvandvas do not exist in Germanic (and in fact she regards dvandvas in Sanskrit as syntactic constructs, not as compounds). Wälchli (2005) notes that dvandvas, or co-compounds using his terminology, are predominantly found in continental Asia, easternmost Europe and New Guinea (see e.g. Nicholas & Joseph 2007 for some discussion on verbal dvandva compounds in Greek) and forms like *Alsace-Lorraine*, *Austria-Switzerland*, *Koptjevskaja-Tamm*, and *whiskey-soda* are not dvandvas but fusional compounds (according to which two entities undergo fusion).[125] Wälchli mentions as exceptional dvandvas in English forms like *today-tomorrow*, *July-August* and *Saturday-Sunday*. We will keep using the term copulative as a cover term, since this seems to be the most widespread use of the term nowadays. In the case of (25), copulative compounds include the two subtypes of compounds: dvandva and appositive, but as will be seen in the following classification, i.e. Plag's (2003), copulative compounds will include appositional and coordinative compounds, and the term dvandva will be used as a synonym of copulative.

Finally, dvandva compounds are also understood differently by Booij (2005: 81), who believes that they always have a plural reading, and Bauer (2003: 43), who understands that the entity denoted by the compound will determine its number (e.g. Tamil *appaa+v+amma* (father+empty.morph+mother) 'parents', Vietnamese *sõt+rét* (be.hot+be.cold) 'malaria')[126]. It is generally agreed that the two elements of dvandvas are in principle interchangeable, but in practice the order is fixed by tradition. (See

---

[124] Bauer (2008) divides dvandvas into five subtypes: additive, co-hyponymic, co-synonymic, compromise and exocentrics. He claims that, in addition to the additive type (e.g. *Alsace-Lorraine*, *Time Warner*), English has compromise compounds like *north-west* and *blue-green*. We treat such data differently (see the subsections of nominal and adjectival compounds in English for our treatment; they are included in (36b) and (55a) respectively).

[125] Following Wälchli, for example, in *Austria-Switzerland* the two countries underwent fusion when they organized the European football championship in 2008 together (not a political fusion), and in *Koptjevskaja-Tamm* there is fusion of the names (not of the persons).

[126] Gloss provided by Bauer (2003).

Benveniste 1974: 147 for other examples of dvandvas, like *pitárā+mātárā* (father+mother), the classic example from Vedic).
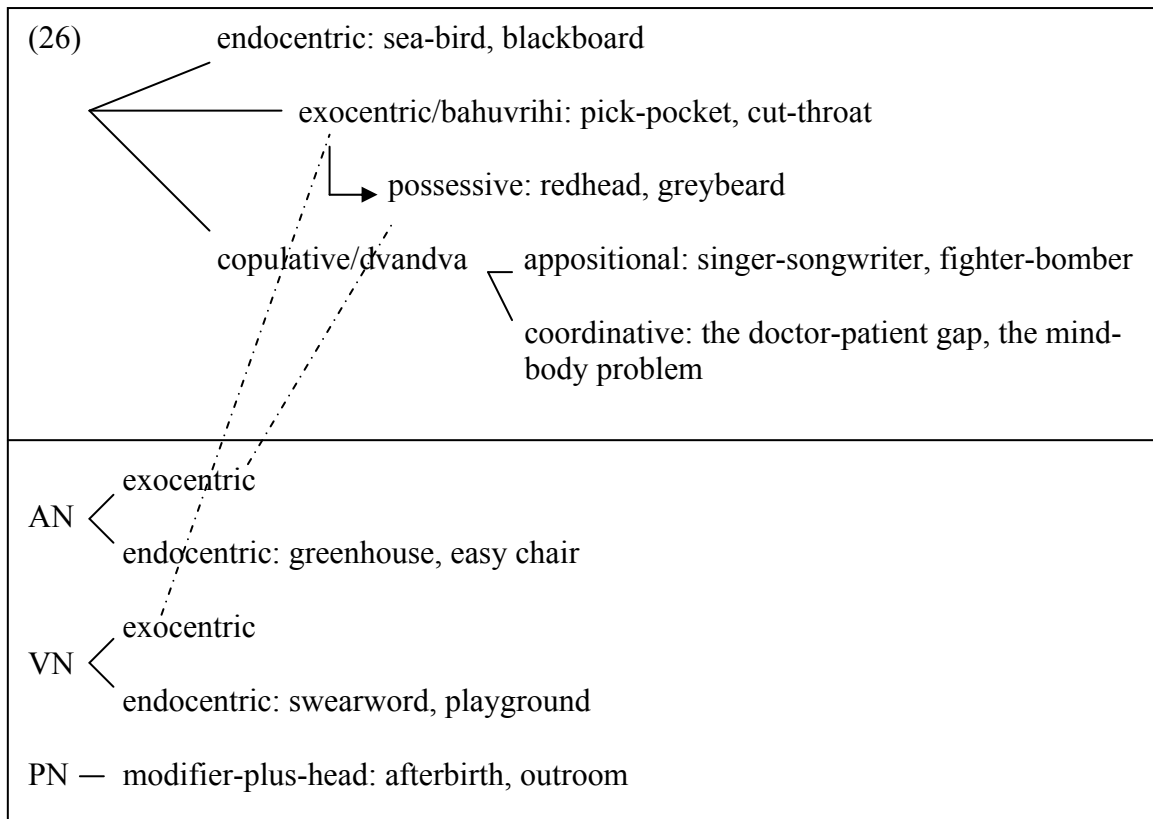
The schematic representation in (25) represents an improvement on the previous semanticocentric approaches to compounding, because it is based on a combination of criteria, but they are not used consistently and uniformly. For example, endocentric and exocentric compounds are defined by being headed or headless, but the same criterion is not applied to copulative compounds, when their two subcategories are precisely distinguished by being headed (appositive) and headless (dvandva). In addition, the head is sometimes interpreted as being the morphosyntactic head[127] and semantic head together but on other occasions it refers to the semantic notion of head only, as is the case for bahuvrihi compounds. That is, bahuvrihi compounds are usually defined by the hyponymy criterion only, and as a result they are said to be exocentric because they are not hyponyms of the head (hyponymy clearly being a semantic notion), although the noun seems to be the morphosyntactic head (cf. Booij 2005: 79-81). That is, some authors are not careful enough to distinguish the two senses of head (morphosyntactic and semantic) and use the term indistinguishably, thus creating confusion and non-uniformity in their analyses.

Although the classification just described is intended to accommodate all types of compounds, it is usually applied to nominal compounds only. Plag (2003) uses a similar scheme for nominal compounds, but he uses as his main criterion the syntactic category of both the input and output categories.[128]

First, Plag describes nominal compounds in general and distinguishes three (or four) types of them (i.e. endocentric, exocentric/bahuvrihi, possessive, and copulative/dvandva), as the first part of the scheme in (26) shows. Afterwards, he focuses on AN, VN and PN compounds, the second part in (26).

---

[127] Recall that the morphosyntactic head gives the syntactic category to the compound as a whole (cf. subsection 2.1.2.1)

[128] See Voyles (1974) for another syntactic-based classification, which divides compounds into derivative compounds and pure compounds. The division seems rather artificial, though: he only mentions that in the first group the first element can be either a preposition or a non-deadjectival adverb without saying anything about the second element, and that in the second group each element can be a N, a V or a deadjectival adverb.

(26)   endocentric: sea-bird, blackboard

      exocentric/bahuvrihi: pick-pocket, cut-throat

        → possessive: redhead, greybeard

    copulative/dvandva appositional: singer-songwriter, fighter-bomber

          coordinative: the doctor-patient gap, the mind-body problem

   exocentric

AN

   endocentric: greenhouse, easy chair

   exocentric

VN

   endocentric: swearword, playground

PN — modifier-plus-head: afterbirth, outroom

Like Bauer (1983, 2003), Plag (2003) does not distinguish the terms copulative and dvandva compounds but he uses them as a cover term for two subtypes, which he calls appositional (*singer-songwriter*) and coordinative (*the doctor-patient gap*). Other authors do not make a distinction between appositional and coordinative compounds because, in their view, all these compounds involve a coordinate relation. For example, Olsen (2001, 2004) only talks about copulative compounds, by which she understands the appositional compounds and the embedded compound in the coordinative compounds in (26). When inserted in the non-head position of a larger compound, the final interpretation of the embedded copulative compound is subject to the semantic requirements imposed by the head. For example, a relational head like *gap* requires the individual components of the embedded compound to stand in a 'between' relation to the head (see sections 2.3 and 2.4 for our proposal concerning copulative compounds).

   Still on the first part of the scheme in (26), exocentric and bahuvrihi compounds are also used as synonymous and Plag uses the term 'possessive' for what are traditionally called bahuvrihi compounds (e.g. *redhead, greybeard*), although, as will be seen in the next paragraph, he is not totally clear about whether they are a subgroup of
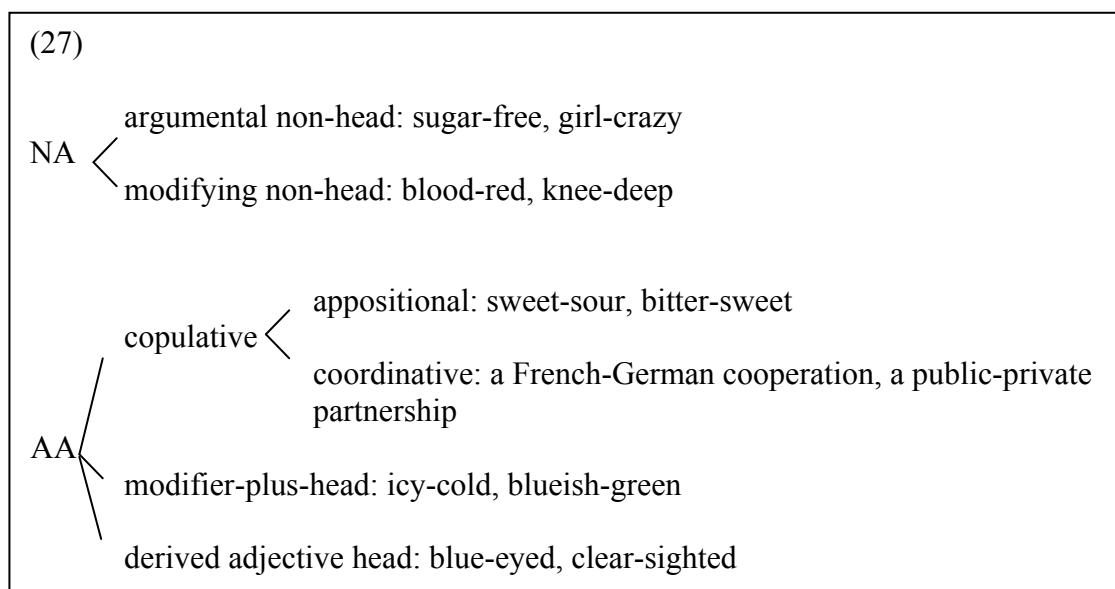
exocentric/bahuvrihi compounds or another type of compound at the same level of endocentric, exocentric/bahuvrihi and copulative/dvandva compounds.

Concerning the general description of nominal compounds (the first half of (26)), Plag's classification basically suffers from the same defects as the previous one. As Plag puts it: "Apart from endocentric, exocentric, and possessive compounds there is another type of compound [copulative/dvandva] (…)" (p. 146). In other words, he puts heterogeneous categories at the same level of analysis. If Plag defines endocentric vs. exocentric compounds as those that have the semantic head inside or outside the compound, possessive compounds are either endocentric or exocentric (exocentric on Plag's view), and not an altogether different type of compound on the same level as exocentric/bahuvrihi compounds, for instance. Similarly, Plag defines copulative/dvandva compounds as those that have two heads, a criterion not used for the other types. That is, presence of a head (endocentricity) or its absence (exocentricity) is considered as a criterion equal to the number of heads present in the compounds, two different criteria not incompatible with each other a priori (see below).

Now, focusing on the second half of (26), we can notice that NN compounds are missing. It is not clear to us whether Plag intended to apply the general description of nominal compounds only to NN compounds or whether he left the reader the task of extracting the NN compounds from the endocentric and copulative/dvandva compounds of the general description. If the former were the case, it would be incongruous because the examples he provides for exocentric/bahuvrihi and possessive compounds are not of the NN type, but rather of the VN and AN type, respectively. If the latter were the case, each type of nominal compound (NN, AN, VN, PN) is defined in rather different terms, an undesirable result. The endocentric vs. exocentric dichotomy is only present in AN and VN compounds, NN and PN being all endocentric according to Plag. Also, the copulative/dvandva compounds are only available to NN compounds; and PN compounds are defined in terms of modifier/modified relationship, not present in any other type of compound. In short, either option of analysing nominal compounds has problems.

We find similar problems of inconsistency in adjectival and verbal compounds. For example, adjectival compounds are divided into NA and AA compounds, the former subdivided into those whose left-hand member is an argument (*sugar-free* 'free of sugar') and those in which it is a modifier (*blood-red* 'red like blood'). Concerning AA compounds, they are subdivided into three subgroups: (i) copulative, which can be
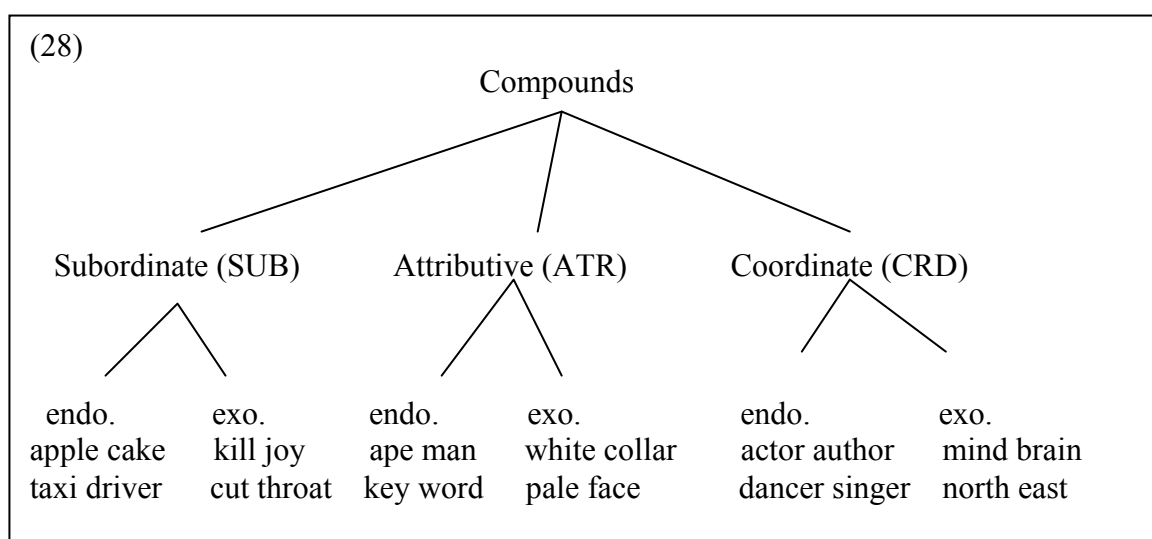
appositional (*sweet-sour*) and coordinative (*a French-German cooperation*), (ii) modifier plus head (*icy-cold*), and (iii) derived adjectives as heads (*blue-eyed*). A schematic representation of adjectival compounds is given in (27). As can be seen, NA compounds are defined uniquely in terms of the argument structure of the head (i.e. whether the N is an argument or a modifier of the head), whereas AA compounds are characterized by a mixture of criteria: by the semantic relation between the two elements in the first two types of compounds (i-ii), and by a formal criterion in the third type (iii) (the head is derived, which is comparable to secondary compounds found in nominal compounds). In addition, the criteria used for distinguishing different types within nominal and adjectival compounds are different.

(27)

NA
- argumental non-head: sugar-free, girl-crazy
- modifying non-head: blood-red, knee-deep

AA
- copulative
  - appositional: sweet-sour, bitter-sweet
  - coordinative: a French-German cooperation, a public-private partnership
- modifier-plus-head: icy-cold, blueish-green
- derived adjective head: blue-eyed, clear-sighted

Finally, verbal compounds are divided into NV (*proof-read*), AV (*deep-fry*) and VV (*stir-fry*). None of the verbal compounds is given a characterization: Plag only calls the last type appositional.

Carstairs-McCarthy (2002) also uses the syntactic category of the input and output categories as the main criterion for the classification, like Plag, along with the notion of head. (The reader is referred to the original works for details of the different classifications briefly reviewed here, and to Bisetto & Scalise 2005 for a review of more traditional classifications like Bloomfield 1933, Bally 1950, and Marchand 1969 and more recent ones like Spencer 1991, Fabb 1998, Olsen 2001, a.o.).

Despite not being consistent, all compound classifications so far have pointed out several criteria and aspects that could be taken into account for a complete classification. The principle to follow looks simple: the more the criteria used for a compound classification, the richer it will be, but in practice, as we have seen, the criteria do not apply systematically to all compounding types, resulting in inconsistency. An attempt to solve this problem from most compound classifications is made by Bisetto & Scalise (B&S, henceforth) (2005: 326),[129] who provide a classificatory scheme for compounding based on consistent criteria and intended to be universal, which is as follows:

(28)

```
                        Compounds
            /               |                \
   Subordinate (SUB)   Attributive (ATR)   Coordinate (CRD)
      /      \            /      \            /        \
  endo.     exo.      endo.     exo.      endo.       exo.
 apple cake kill joy  ape man  white collar actor author mind brain
 taxi driver cut throat key word pale face  dancer singer north east
```

The main criterion of analysis is the type of grammatical relation holding between the two elements of the compound. By this criterion, B&S identify three macro-types, each defined by a different relation. One of them is a relation of complementation, which can be found in compounds like *car-driver*, where *car* is understood as the internal argument of *drive*, *book cover*, interpreted as the 'cover of a book', and *catfood*, understood as 'food for cats'. This complementation relation gives rise to SUB compounds, which are contrasted with the two other macro-types: ATR compounds and CRD compounds. The former are characterized by a modification relation: the first

---

[129] B&S's (2005) classification is revised in Scalise & Bisetto (2009). Since there are no substantial changes, the discussion to follow is based on B&S (2005). Only one difference between the two works will be pointed out: subordinate compounds are divided into ground and verbal-nexus compounds in Scalise & Bisetto (2009). 'Ground' compounds correspond to 'primary compounds' but 'verbal-nexus' compounds are not identical to 'secondary' compounds. The difference lies in the fact that the relation between the underlying verb in the deverbal head and the non-head is different in the two types: the non-head can be an argument or an adjunct (e.g. *bookseller*, *street seller*) in the case of verbal-nexus compounds, but only an argument in the case of secondary compounds.

element is a property attributed to the second element, as in *blue cheese* and *pale face*. The latter macro-type is defined by a coordinating relation, 'and' being the typical conjunction found in Indo-European languages, as in *poet-painter*.[130]

These three macro-types are in turn defined by a second criterion: the presence or absence of a head (endocentric vs. exocentric), which divides each macro-type into two sub-types.

B&S's first level of analysis (i.e. the grammatical relation between the two elements of the compound) may give categories which are too roughly defined, but homogeneity is achieved by this criterion, and further distinctions in each macro-type are not denied (p. 331):

(29)   "The proposed classification considers a first level of analysis, that is the grammatical relationship between the two constituents. We do believe that this first step is basic and that it should be kept separated from other possible criteria such as the internal structure, the semantic relation between the constituents, the origin of compound constituents or the categorial status of the constituents; all these criteria have to be ordered, so to speak, after the grammatical level of classification."

After outlining B&S's classification, let us consider it in more depth by looking at several points, some of which can be seen as strengths and others as weaknesses of the compounding scheme.

First, B&S support their tripartite division of compounding types with the fact that each type has a different selection mechanism. That is, the head of the compound selects the non-head differently in each of the three macro-types. B&S represent compounds in Lieber's (2003, 2004) framework of lexical semantics, according to which each lexeme is characterized by a skeleton, which contains syntactically relevant information, and a body, which is the encyclopaedic, holistic and idiosyncratic

---

[130] Adopting B&S's classification, Lieber (2008) believes that ATR compounds are the default class where compounds without an argumental or a coordinate relationship are placed. In this respect, note that CRD and SUB compounds can have a secondary ATR interpretation. For example, *mother-child* can be interpreted as a child who plays being the mother, in addition to the CRD interpretation (i.e. the relation between mother and child). Note that the CRD interpretation arises only with the presence of a noun like *relation* outside *mother-child* (an observation which will become relevant when discussing the so-called CRD compounds).

information, variable from speaker to speaker.[131] Now let us see how selection works in each sub-type of compounding following Lieber's system. In CRD compounds like *actor-director*, the skeletal features of the two constituents are identical and the body features are almost the same. As for SUB compounds like *apple cake*, the skeleton is irrelevant, and as for the body features, at least one of them must be matched by the two elements of the compound: <edible> could be the matching feature between *apple* and *cake*.[132] Finally, in ATR compounds like *snail mail*, the skeleton seems irrelevant and again at least one body feature of the two constituents must be matched. In the case of *snail mail*, *mail* has the encyclopaedic feature <takes time> which can be matched with the encyclopaedic feature <very slow> of *snail*, all other body features of the non-head being invisible to the head, e.g. the fact that snail <secretes slime>. (See B&S 2005: 329-330 for details and for their representation of each compound in Lieber's framework).

These three different selection mechanisms are meant to reinforce the tripartite classification but are not enough to account for all compounds. Scalise et al. (2005) need to introduce two subcategories into SUB compounds: primary and secondary compounds, because the selection works differently in each subcategory. The mechanism of selection in primary compounds works in the way described for *apple cake*, but in the case of secondary compounds, selection is mainly carried out by means of the skeleton, and the body features are not so relevant (if at all). That is, the non-head satisfies an argument of the verbal base in head position, as in *car driver*. In short, there are four types of selection mechanisms for three big types of compounds, not the ideal situation if the selection mechanisms were intended to give support to the initial classification. In addition, the difference between SUB primary compounds and ATR compounds in terms of selection is minimal. In both, the skeleton does not seem to play a role and in both at least one body feature of the head needs to match one body feature of the non-head. The difference between the two types of compounds is reduced to the

---

[131] Lieber's skeleton is comparable to Jackendoff's (1990), Rappaport Hovav & Levin's (1998), and Levin & Rappaport Hovav's (2005) level of lexical conceptual structure while Lieber's body can be compared to the constant/root in Rappaport Hovav & Levin's and Levin & Rappaport Hovav's terms.

[132] One might argue that cakes do not necessarily have the body feature <edible> in compounds like *glue cake*. What is relevant, though, is that given a plausible context, there must be a matching feature in the body features of the two compounding elements. If that is the case, the compound is allowed. In the case of *glue cake* the matching feature could be 'ingredients', i.e. <made with ingredients> could be a body feature of *cake* and <can be an ingredient> could be a body feature of *glue*. The context could be a competition in which participants had to make a cake with unusual ingredients, and the result was a glue cake, a sand cake, etc. A different question, which is not addressed by B&S, is why the skeleton and body features are relevant in some cases and not in others.

fact that in the case of ATR compounds "What matters is that the non-head matches at least one of the encyclopaedic features of the head." (p. 330) and in the case of SUB compounds "At least one of the features of the head constituent must be matched by the encyclopaedic features characterizing the non-head constituent." (p. 330). To our understanding, the difference between the two types of compounds is null, or at most can be reduced to one of directionality, i.e. whether the need for the matching feature comes from the head or the non-head. The distance separating two macro-types of compounds, namely SUB primary compounds and ATR compounds, is then shorter than two subtypes belonging to the same macro-type, namely SUB primary and SUB secondary compounds, an undesirable result as far as we can see. In short, the selection mechanisms found in compounding, as discussed by B&S, do not support their three macro-types, as it was originally intended.

A second point (although a minor one in this case) which should be clarified in B&S's classification is the fact that, although they admit the possibility of further subdivisions into the three big types of compounds and they in fact propose to divide SUB compounds into primary and secondary, as we have just seen (cf. Scalise et al. 2005), they do not mention how these two subtypes interact with the second level of analysis in their classification (endocentricity vs. exocentricity, but see Scalise & Bisetto 2009 where this interaction is made explicit), and with the other two macro-types (ATR and CRD compounds) if there is some interaction. In fact, the primary/secondary distinction is distributed unevenly among the three macro-types in English. ATR and CRD compounds are all primary compounds and only a subset of SUB endocentric compounds are secondary compounds. Recall (from subsection 2.2.2) that secondary compounds are understood as having a (de)verbal head and a nonhead functioning as an argument of the head (e.g. *truck driver*) and primary compounds are, by contrast, compounds whose head is not (de)verbal or the nonhead is not interpreted as an argument of the (de)verbal head (e.g. *greenhouse*, *mass production*). Such criticism against non-homogeneity cannot be levelled against B&S, because they are aware that homogeneity may not exist beyond the first level of analysis: "And exactly the grammatical relations between the constituents of a compound can allow a homogeneous grouping (at least on a first level) of compounds of different languages" (p. 331), but we believe that the interaction of finer subdivisions (primary vs. secondary) with the two levels of analysis should be made explicit. Although

homogeneity is a desirable property in every level of analysis, it remains to be seen whether it is a real possibility.

Another point which is not totally clear to us from B&S's classification is the way they understand their second level of analysis: endocentricity vs. exocentricity, a point which is revised in a later work. Scalise & Guevara (2006) acknowledge that "presence or absence of a lexical head" can be ambiguous between a formal head and a semantic head, and they define each type as follows (p. 190):

(30) "The formal head of a compound is the constituent which shares with –and percolates to- the whole compound all of its formal features: lexical category and subcategorization frame. The whole compound, thus, is expected to have the same distributional properties of its formal head."

"The semantic head of a compound is the constituent which shares with –and percolates to- the whole compound all of its lexical-conceptual information (LCS in short, following Jackendoff 1990 and Lieber 2004). The whole compound, thus, is expected to be a hyponym of its semantic head."

Their claim is that endocentricity refers to those compounds where the formal head and semantic head coincide, as in *capo+stazione* (lit. master+station, 'station master') in which the semantic head (a *capo*, which is a hyperonym of a *capostazione*) is the same as the formal head (the masculine gender of the compound comes from *capo*: $[[capo]_{masc}[stazione]_{fem}]_{masc}$).[133] When the two heads do not coincide, then the compound is exocentric. In their terms (p. 192):

(31) "An *endocentric compound* has at least one formal head and at least one semantic head. If a compound has only one formal head and only one semantic head, then the two must coincide.

If a compound realises any of the remaining possibilities, it will be considered to be *exocentric*."

---

[133] From such an example, we notice that Scalise & Guevara (2006) understand 'subcategorization frame' as including, at least, features like gender (e.g. masculine vs. feminine gender in Romance), and plurality although they do not mention the latter explicitly. In addition to syntactic category, gender, plural and tense marking will become relevant in determining the formal head in English and Catalan compounds (cf. subsections 2.3.1. and 2.3.2).

The improved understanding of the notion of head led Scalise & Guevara (2006: 191) to change their position (Scalise et al. 2005) concerning the status of CRD compounds. First, they claimed that CRD compounds, despite the two elements of the compound having inflection in languages like Italian, had only one head, which was determined by the constituent occurring in the canonical head position in the language: Italian would have the head on the left, e.g. *bar pasticceria* 'bar-pastry shop', and English would have the head on the right, e.g. *actor-manager*. Such a view is changed to incorporate the two notions of head (cf. 30, 31), and accordingly CRD compounds are now claimed to have two heads.

To sum up, we have seen that, of the three signs of weakness that could initially be attributed to B&S's classification, two of them (i.e. finer subdivisions not interacting with their two levels of analysis and the initially unclear notion of head) are not real problems, but the third one is. That is, the head selection mechanisms in compounding as discussed by B&S do not match their three macro-types of compounds. Despite this fact, their classification is the most promising classification of compounds available. It is intended to be universal and its first criterion of analysis is applied consistently and gives three homogeneous macro-types of compounds, which are in turn subdivided into two subtypes each: endocentric vs. exocentric. We have also seen that more work needs to be done to make finer subdivisions within each subtype, also aiming at homogeneity ideally. It is partially our purpose to carry out such a task in what remains of this chapter. In other words, in the next section we will present an exhaustive study of the compounds available in English and Catalan using the categorial status of the input and output categories as a main criterion for the classification. Such a procedure is justified in that it will allow us to come up with a more refined compounding classification and to corroborate in chapter 3 Snyder's (2001) hypothesis about the alleged correlation between productive compounding and resultatives in a language. While presenting the compounds of English and Catalan according to their syntactic category (input and output), some comments will be made within each categorial type to better understand the relation between the two elements of the compound, comments that could also be used to add further subdivisions to the ones we intend to make here but that we leave for future research. Instead, we will limit ourselves to the task of incorporating the categorial subtypes into the original compounding scheme of B&S, thus creating further homogeneous subdivisions in their classification. Although B&S's classification will be our starting point and, as already noted, compounds will initially be classified according

to their compounding scheme, we will argue for a unique compounding type: our proposal will be that all compounds are based on a head vs. non-head relation, from which the different interpretations arise (subordinate, attributive). The existence of CRD compounds and exocentric compounds will be put into question. Our proposal will be gradually introduced in the following section, after considering how B&S would classify English and Catalan compounds, and further developed in the discussion section.

This concludes our general (and sketchy) survey of compound classifications. In short, classifying compounds has proved to be rather difficult, and, as we have seen, it has been carried out in a number of ways: (a) by allegedly underlying syntactic phrases from which the compound is derived (e.g. Lees 1960), (b) by semantic classes (e.g. Hatcher 1960, Levi 1978, Downing 1977), (c) by the syntactic categories of the lexemes that make up the compound (e.g. Plag 2003, Carstairs-McCarthy 2002), and (d) by a mixture of the previous methods (e.g. Bauer 1983, 2003, Booij 2005, Adams 1973), among others. We have concluded that any classificatory scheme is bound to be controversial if the criteria used for the classification are not used consistently. B&S's (2005) classification seems to precisely overcome this problem but, as will be seen later, it presents some other problems when considered thoroughly (that are also present in other studies of compounds).

2.3 English and Catalan compounding

In the classifications of English and Catalan compounds to follow, the use of clipping, blends, and reduplication in compound formation will not be taken into account. Also, compounds having more than two constituents will be omitted as well as neo-classical formations and phrasal compounds. Finally, discussion about the presence of linking elements will be absent unless they have some effect on the categorial status of the compounding elements.

There is no need to look at clipped compounds because they will not provide us with new data, given that they are based on unclipped compounds, and there is no change in input and output category. The only difference between clipped and unclipped compounds is a stylistic one. Some authors such as Booij (2005) have claimed that only one of the members can be clipped, as in *e-mail*, German *U-Bahn* (*Untergrund-Bahn*

'metro'), but there are some data showing that both elements can be clipped: e.g. *sci-fi* (*<science fiction*). Some Catalan examples of clipping are more minimalist: of the two compounding elements, only one remains, as in *auto(mòbil)* 'automobile' and *tele(visió)* 'television'. (For similar examples in Spanish, see Buenafuentes 2007: 391-400). Note, though, that they are cases of neo-classical compounds. Clipping is predominantly found in technical language.

Blends share with compounds the fact that they involve two lexemes. Some examples are *brunch* (*<breakfast+lunch*), *glasphalt* (*<glass+asphalt*), *wargasm* (*<war+orgasm*) and *paraloon* (*<parachute+balloon*). Like clippings, blends cannot lead us to new conclusions different from the ones which can be drawn from the source words on which the blends are based. So, blends will not be taken into consideration in the discussion of English and Catalan compounds. (See e.g. Olsen 2001: 901 for recent blends and Adams 1973 for a classification of blends into syntactic (e.g. Verb-Object like *bus-napper* (*<bus+kidnapper*)) and semantic types (e.g. Locative, like *chunnel* (*<channel+tunnel*))). Catalan has borrowed some blends from English like *motel* and *aparthotel*. Other examples are *cantautor* (*<cantar+autor* sing+author) 'singer-songwriter' and *autobús* (*<automòbil+òmnibus* automobile+bus) 'bus'.

Reduplication refers to the word-formation process, according to which some part of the base (or the entire base) is used more than once in a word (Bauer 1983: 212-213, 2003: 31-32, Adams 2001: 127-129). They are a special kind of compound in that they are phonologically motivated. Some examples from English include *hocus-pocus*, *teeny-weeny, tick-tock*, *chitchat*, and *pitter-patter*. From Catalan, there are examples like *ning-nang* (the sound made by a bell), *pengim-penjam* (person clothed in an untidy way), *piu-piu* (noise made by poultry), *xiu-xiu* (whispering), and *ziga-zaga* (zig zag). Note that sometimes one of the elements does not exist independently in the language. In fact, reduplicative formations are sometimes classed as compounds to avoid calling them affixes, which would be an odd position since one word would be made up of two affixes (cf. Katamba & Stonham 2006). However, for their peculiarities they will not be taken into account in the computation of English and Catalan compounding.

Some authors do not view clipped forms, blends and reduplicative formations as proper compounds, a further reason not to consider them here (see Kubozono 1990, in Adams 2001, for clippings; Plag 2003 for blends, and Katamba & Stonham 2006 for reduplicative forms)

We will not look at compounds containing more than two elements either, given that they are ultimately based on binary structures (e.g. [[*pillow cover*] *painter*] and [*neteja* [*para brises*]] (clean stop breezes) 'windscreen cleaner'), and hence cannot give us new insights into the patterns of compounding available (leaving recursivity aside for the moment). Also, as already noted earlier, neo-classical compounds and phrasal compounds will not be looked upon either. The former are characterized by a number of distinct properties that distinguish them from ordinary compounds, and their formation seems to be subject to a different process (cf. subsection 2.2.1: 102-104). The latter can superficially be taken as non-phrasal compounds. Recall that any phrase inside a compound is treated as a simplex word: [phrase]$_{NP, PP, VP, CP}$ → [phrase]$_{X^o}$. That is, when the phrase is incorporated into the compound, the result behaves like an ordinary compound: [[phrase]$_{X^o}$ + X^o] (cf. subsection 2.2.1: 104-109). Their omission is then explained for expository reasons (but see our treatment of CRD compounds). Finally, the presence of linking elements such as case, plural markers, possessive *'s* and linking vowels are not dealt with (e.g. *fees controversies, women's magazine*), because they do not have any repercussion on the syntactic categories of the constituents of the compound, which is our main point of interest. Only some remarks are made about the presence of a linking vowel in a compound type found in Spanish, the N+i+A type (but not in its Catalan counterpart). (For an active role of linking elements in Germanic and Romance compounding, see e.g. Delfitto & Melloni 2008).

The following two subsections (subsection 2.3.1 and subsection 2.3.2) present an inventory of the compound types available in English and Catalan. The input and output category is used as the main criterion for the survey, a criterion which is applied to B&S's (2005) classificatory scheme, thus adding finer subdivisions into their original scheme. The use of the input and output categories may seem problematic because many English words can belong to more than one category without any change in form. The existence of a compound type, though, does not depend exclusively on a single ambiguous compound and hence can be corroborated by looking at whether there exist compounds of the same type without ambiguous categorial members. In addition to syntactic category, some additional information is provided to better understand the semantics involved in the compound. After presenting the compounds according to B&S's classification with our additional level of input/output category, some remarks will be made about some compound types, especially the CRD ones and exocentric ones, whose existence will be questioned. The compounds provided are taken from a

number of sources (classical and more recent contributions) in order to really present the compound types available nowadays. Note that the original form of a compound is kept the same as in its source, with the consequence that some compounds are written as one word, others are hyphenated and still others are spelt as two separate words (cf. footnote 4 in chapter 1).

2.3.1 English

Before delving into the English survey of compounds, this subsection presents some convenient introductory remarks. First, some sources on which the survey is based are presented. Second, the stress criterion for identifying compounds is briefly discussed, which is followed by other tests for distinguishing compounds from phrases. Finally, the existence of prepositional compounds is questioned.

Sources for the survey of English compounds include Ackema (1999a, b), A&N (2004), Adams (1973, 2001), Allen (1978), Bauer (1983, 2001, 2003, 2004), Bauer & Renouf (2001), Berg (1998), Bloomfield (1933), Booij (2005), Carstairs-McCarthy (2002), Dijk (1997), Downing (1977), Katamba & Stonham (2006), Lees (1960), Levi (1978), Lieber (1983, 2003, 2004, 2008), Marchand (1969), Matthews (1974, 1991), McIntyre (2009), Meyer (1993), Plag (2003), Selkirk (1982), Spencer (1991, 2000, 2003b), and Wälchli (2005).

Compounds (as opposed to phrases) have traditionally been identified by the presence of stress on the first constituent, which has come to be known as the stress criterion. The literature on stress as applied to compounding (especially NN compounding) is vast (Bauer 1998, 2004, Chomsky & Halle 1968, Cinque 1993, Giegerich 2004, Ladd 1984, Liberman & Sproat 1992, Olsen 2000a, Spencer 2003b, among many others) and the stress criterion has proved to be rather controversial. For example, Bauer (2004) concludes that the function of stress is not to identify a compound but to indicate contrast, lexicalization, and has a naming (vs. descriptive) function. Spencer (2003b) also agrees that stress does not give clear results as to what is a compound, and observes that stress may be associated with lexicalization rather than compoundhood. In view of the fact that we find compounds with compound stress (*'blackbird, 'cheese cake*) but also compounds with phrasal stress (*Compound 'Stress, apple 'pie*), we conclude that the traditional test which identifies compounds by having left-stress and phrases by having right-stress seems rather difficult to maintain, and hence will not be taken into account in what follows (but see Giegerich 2004, 2005).

There are other criteria which have been used to identify compounds (vs. phrases) in English, some of which follow (see e.g. Giegerich 2004, 2005, Lieber & Štekauer 2009, McIntyre 2009). Neither compounding element can be independently modified, i.e. the two elements are inseparable, (e.g. *coffee (*big) cup* vs. *morning hot coffee*). Further, neither compounding element can be replaced by *one* (e.g. *the tea drinker (*and the coffee one)* vs. *the city employee and the state one*) nor can they undergo deletion in coordination. The last test has proved to be controversial: it has been suggested that deletion in coordination is neither a test for a lexical process (i.e. compounding is a lexical process according to some authors) nor a test for a syntactic process. Rather, a phonological constraint seems to underlie deletion in coordination, according to which the deleted part must at least constitute a phonological word (e.g. *clock and watch-maker* vs. **kind- and happily*) (see Booij 1985). In short, it seems that the inseparability test and the *one*-replacement test are the most reliable tests to identify compoundhood (vs. phrasehood) and, as a result, will be used in what follows whenever the nature of a sequence of elements needs disambiguation.

As for the types of English compounds, the majority of discussions about English compounding do not include prepositional compounds. Examples like *into*, *onto*, *upon*, *without* and *within* which could instantiate prepositional compounds are lexicalizations of two prepositions frequently occurring together, which have developed a unitary semantic interpretation with the consequence that they are perceived as one word by speakers. In addition, new formations based on the P+P pattern appear to be impossible: **withby*, **upunder*. However, forms like *outdoors*, *offstage*, *overhead*, *uphill* and *underfoot*, which are the union of a preposition and a noun, could be seen as prepositional compounds. This is the position defended, for example, by Boertien (1997) but the speakers consulted do not agree on the productivity of such forms, which explains why we leave them out from the present survey of English compounding (but we hope to study them further in future research). In the subsections of nominal and adjectival compounds, we will see that these forms can act as nouns and adjectives.

Our survey of compounding in English starts with nominal compounds (subsection 2.3.1.1). Then verbal compounds are presented in subsection 2.3.1.2, and finally adjectival compounds are discussed in subsection 2.3.1.3.

2.3.1.1 Nominal compounds

[NN]_N compounds are the most productive type of compounds in English but they are not the only nominal compounds, although Spencer (2003b: 330) observes that "'true' compounding can only refer to NN collocations", arguing that [AN]_N compounds, for example, are lexicalized phrases. We will see below that not all cases of [AN]_N compounds are lexicalizations of phrases and that there are other types of nominal compounds as well, although it is true that they may not be as productive as [NN]_N compounds.

This subsection includes the different types of nominal compounds: first the compounds whose second member is a noun are presented (NN, VN, AN, PN), which are followed by those which do not conform to this formal criterion (NA). Then, some discussion about the status of three different types of formations (VV, VP, PV) is provided. Finally, a table summarizes the results of this subsection.

[NN]_N compounds[134]

This is the most common type of compound. In [NN]_N compounds, it is generally assumed that the role of the first noun is to make the meaning of the second noun more precise, as in *olive oil* and *paper clip*. If this restriction holds, two predictions follow: the first noun cannot denote a superset of the head noun or a necessary part of it (cf. Meyer 1993). So, compounds like *food chocolate* and *leg trouser* should be ungrammatical with the intended (uninformative) meaning and they are indeed, which explains the absence of such compounds in the examples below.[135]

Following B&S's (2005) classification, [NN]_N compounds can be divided into SUB (32), ATR (35) and CRD (36) compounds, each of which will be dealt with in turn. All SUB compounds are endocentric and three subdivisions can be made depending on the nature of the head: (a) whether it is a relational noun, (b) whether it is a deverbal noun, and forms a secondary compound with the first constituent, or (c) whether it is neither a relational noun nor a deverbal noun.

---

[134] It is not totally clear whether N's N constructions (e.g. *shepherd's pie, driver's seat, women's magazine*) are compounds (e.g. Lees 1960, and Taylor 1996, cited in Adams 2001) or not (e.g. Bauer 2001, Spencer 2003b), which explains their omission here. Note that if they turn out to be compounds, the possessive *'s* does not change the category of either constituent of the compound, hence this type of compound could be assimilated to NN compounding in this respect.

[135] Of course the compounds are grammatical with the non-intended meaning, such as 'chocolate that I always eat with other food, as opposed to chocolate that I eat on its own' in the case of *food chocolate*.

(32) a. animal doctor, arrowhead, bedside, bootleg, bottleneck, brain death, brain surgery, car thief, car mechanic, catgut, cookbook author, crew member, finger surgery, fingertip, horse doctor, masthead, pinhead, probation officer, roads lobby, roadside, sea surface, silk merchant, **table leg**, and tooth decay.

b. anteater, bear-baiting, beer-drinker, bicycle-repairer, bicycle-repairing, **bookseller**, brick-layer, brick-laying, cake baker, church-goer, coffee-maker, consumer protection, crime prevention, dish-washer, gamekeeper, globe-trotter, grave-digger, hair restorer, hay-making, heart-breaker, heart-failure, life-insurance, mail delivery, money-changer, moneylender, nutcracker, pasta-eater, pasta-eating, population growth, potato-picking, sheep-shearing, shoemaker, shop clearance, sign writer, slum clearance, soccer-playing, souvenir-hunting, stage manager, story-teller, sun-worshipper, sword-swallower, task assignment, tax-evasion, time-saver, tongue-twister, trash removal, truck-driver, typesetter, whiskey-drinker, window-shopping, and wish-fulfilment.[136]

c. advice centre, amusement park, apron string, armchair, banana oil, bar code, bath towels, battlefield, bedtime, beehive, bee sting, body jewel, bookcase, book cover, broomstick, bull ring, bungee-jumping, butterfly net, cable television, cane sugar, car factory, carving knife, chewing gum, chicken fat, Chomsky hierarchy, city wall, cleaning lady, **clog dance**, clothes cupboard, coffee-table, computer desk, computer games, computer surgery, correspondence course, detention centre, dog house, domino theory, doorknob, drinking water, fan dancer, fees controversy, field mouse, film festival, film society, film industry, fish cake, fish farm, flour mill, fruit cake, fruit market, garden-party, gas mask, goat cheese, hairbrush, honey bee, horror film, horse shoe, ice-pack, ignition key, impulse buying, India-rubber, ironing board, language laboratory, laser

---

[136] Occasionally two types of secondary compounds are distinguished: syntactic compounds and synthetic compounds (for the lack of a better term). The former are process nominalizations: they are semantically predictable in that they are understood as object plus verb, can be modified by modifiers which act like adverbs (e.g. *frequent bicycle-repairing*), and cannot be pluralized, among other properties, which contrast with synthetic compounds like *bicycle-repairer*, which are not process nominalizations: they are not semantically predictable (a *drug-user* is a person who uses drugs, but a *nutcracker* is a tool for cracking (the shells of) nuts), cannot be modified by modifiers like *frequent* with the same meaning (e.g. *a frequent bicycle-repairer*), and can be pluralized (e.g. *two bicycle-repairers*) (cf. Oshita 1994, Adams 2001). This distinction will not be made in the present work, since it is orthogonal to the present discussion, although it is an interesting point to pursue in future research.

printer, letter head, life boat, living-room, lodging house, love potion, maple syrup, mass production, metal worker, midnight sun, mincemeat, mine worker, mosquito net, mouse trap, navigation aid, needle work, night flying, observation post, olive oil, paper clip, party drinker, peanut butter, picture book, pitchfork, protection money, prose poem, raincoat, reading glasses, reading material, reception committee, refugee camp, retirement age, rocking horse, safety belt, sand castle, sea bird, sheep dog, skyline, smoke screen, snake bite, space station, spring-cleaning, state archive, steam iron, street seller, student loan, suggestions box, sunburn, Sunday driver, swimming pool, tax law, tea-room, tear gas, tiepin, toothache, traffic lights, Universities yearbook, water-skier, water-skiing, weapons system, wedding dress, Wellington airport, wildlife sanctuary, windshield, windmill, wind storm, and zoo animal.

The endocentricity of the compounds in (32a-c) comes from the fact that the second noun is both the semantic and formal head: for example, a *clog dance* (32c) is a type of dance; and *dance* is also the noun which inflects for plurality. The SUB relation between the two constituents of the compound is evident when the head is a relational noun or a (de)verbal noun: the non-head is understood as the complement of the head, e.g. 'the leg of the table' in *table leg* (32a) and 'the seller of books' in *bookseller* (32b). Concerning the compounds in (32c), the non-head is also subordinated to the head although it is not interpreted as the internal argument. Since the head is neither a relational noun nor a deverbal noun which takes an argument as its non-head, we cannot interpret them by means of purely linguistic knowledge. Their interpretation is rather based on the possible links between the two elements, along with the surrounding discourse and our knowledge of the world, which also play a role in getting a more precise meaning (e.g. a *clog dance* can refer to a dance where dancers have clogs on, a dance where clogs are placed on the stage, a dance where dancers give clogs to the audience, etc.).

Whether the compounds listed in (32b) are NN compounds structurally is not so obvious. There are at least two possible analyses for them. One of them is indeed to treat them as NN compounds, as in (33a), and a second analysis is to treat them as NV compounds, followed by suffixation, as in (33b):

(33)   a. [[*book*]$_N$ *seller*$_N$]$_N$

         b. [[*book*$_N$ *sell*$_V$]$_V$ *er*]$_N$

At first sight the structure in (33a) seems more appealing, given that the process of NN compounding is already available in the language (and *book* and *seller* exist independently in the language) and verbal compounds of the NV type in (33b) are not productive in English (i.e. NV compounds in which the noun is the internal argument of the verb appear to be *systematically* absent). The fact that *book* is an argument of the verb *sell* can be explained by assuming that the deverbal noun *seller* keeps the same argument structure of the base. On the other hand, the non-existence of argumental NV compounds, i.e. *to booksell*, (see the subsection of 'verbal compounds' for other NV compounds that do exist) is not a problem if one allows for an overgenerating morphology (Allen 1978) and, in fact, argumental NV compounding is not totally excluded from Germanic languages, as can be seen by looking at Frisian (cf. Dijk 1997). In addition, idiomatic readings are lost under inheritance (34b) but can be preserved if the noun and verb merge directly, as in (34a). Given that (34c) is grammatical and has an idiomatic reading, we can conclude that the noun and verb must merge directly, i.e. *to troublemake*, favouring the structure in (33b) (cf. A&N 2004: 54-59 for this example and for other arguments which support the structure in (33b); see also A&N 2008).

(34)   a. John always makes trouble.

         b. #John is a maker of trouble.

         c. John is a real troublemaker.

Although we have argued for an NV structure for the compounds in (32b) underlyingly, on the surface they are two nominal word-forms/orthographic words (cf. recall that 'word-form/orthographic word' is one of Matthews's 1974 senses of the term 'word') and this is the view taken by Snyder (2001) when he formulates his Compounding Parameter. In view of examining his hypothesis in the following chapter, we will take the superficial view that the compounds under study are NN compounds, which

explains why they are listed in this subsection.[137] The same reasoning applies to the other compound types.

Regarding ATR compounds (35), they are divided into two subgroups. The compounds of the first subgroup (35a) would be classified as exocentric in B&S's scheme. They treat, as exocentric ATR compounds, [AN]$_N$ compounds like *longlegs* and *paleface* (discussed below: (38b)), which can be compared to the compounds in (35a). Note that such treatment is not congruous: the ATR relation is established on the basis of the two visible compounding elements (the face is pale in *paleface*) but the exocentricity of the compound is based on the compound as a whole and the entity the compound refers to, which is located outside the compound (*paleface* refers to a person whose face is pale). Put differently, the ATR relation and the exocentricity of the compound are based on different elements. If one wanted to maintain the exocentricity of the compound, the relationship between the putative head which lies outside the compound and the compound as a whole would be a SUB relationship: a *sabertooth* is an animal that has saber teeth. However, we think that such treatment is erroneous.

We maintain that the compounds are ATR but we argue for their endocentricity: although it has to be understood metonymically, the second noun is the head semantically. The metonymic extension of the second noun can refer to people (e.g. *skinhead*), animals (e.g. *sabertooh*) and objects (e.g. *hatchback*). For example, *head* stands for 'person' in *skinhead* so that *skinhead* refers to a kind of person (with the consequence that *head* in *skinhead* is the semantic head). The plural form *skinheads* also suggests that *head* is the formal head, given that the plural marker seems to attach to it. However, there are some cases which seem to question the plausibility of treating the second noun as the formal head. Consider *sabertooth*: the head cannot be *tooth* because the plural of *sabertooth* is not *saberteeth*, as one would expect if *tooth* were the head, but *sabertooths*. Recall our suggestion that *sabertooth* is probably seen as a simplex word, and consequently *sabertooth* takes the regular plural marking –*s*. Another apparent problematic case is *butterfingers*, which has a plural ending and can indistinguishably refer to one person and to more than one. The problem of identifying *fingers* as the formal head is as follows: since one person typically has more than one

---

[137] In a construction-based approach to morphology, these compounds would also be regarded as NN compounds with the first N being treated as a modifier of the second one (cf. Goldberg 1995). In a syntactic-based approach to word formation, Borer (2008) also treats the non-head of secondary compounds in Hebrew as a modifier, and not as an argument, of the head.

finger, the word will always be in the plural independently of whether it refers to one person or to more than one. When the compound is meant to refer to more than one person, the plural marker signalling the plurality of people will be indistinguishable from the plural marker signalling the plurality of their fingers. The fact that the two plural markers are fused into one has the consequence that they are not visible separately but it does not mean that *fingers* cannot be taken as the formal head. To recap, the compounds in (35a) are better analysed as endocentric ATR compounds: in *blockhead*, the *head* is like a *block* (metaphorically, hence an ATR compound) and metonymically it refers to a type of person (semantic head). The noun *head* also provides the compound with the nominal category and is marked for plurality (formal head). The compounding pattern of (35a) is regarded as unproductive (Giegerich 2004: 3).

Concerning the compounds of the second subgroup (35b), they are all endocentric ATR compounds and there is no other possible analysis. For example, *jar* is both the semantic and formal head in *bell jar*: the compound denotes a type of jar, one that resembles a bell, and the formal features also attach to *jar*: *one bell jar* vs. *two bell jars*.

(35)   a. birdbrain, blockhead, bonehead, **butterfingers**, cauliflower ears, cottontail, eagle-eyes, egghead, hatchback, pronghorn, razorback, sabertooth, skinhead, and spoonbill.

b. **bell jar**, box kite, bulldog, carrier bag, chain reaction, codfish, crocodile tears, death penalty, demon barber, father-figure, fossil fuel, founder member, football game, handlebar, houseboat, killer virus, mackerel sky, murder charge, prison camp, soldier ant, sponge cake, tenant farmer, and zebra crossing.

Following B&S's classification, the forms in (36) are CRD compounds (the more general term that includes both appositional and coordinative compounds in Plag's scheme in (26)), which can be divided into two groups: (36a) and (36b). The former are endocentric (appositional) compounds while the latter are exocentric (coordinative) compounds. Meyer (1993) remarks that CRD compounds are productive in German despite the restriction that the two nouns must be of the same ontological type, as in his

German examples *Theatermuseum* (theatre+museum) and *Dichterfreund* (poet+friend) where the two nouns denote buildings in the first compound and people in the second one. Such restriction is also observed in the examples listed below. Recall that Olsen (2000b, 2001) also observes that CRD compounds are a productive pattern in English. Notice that Adams (2001: 82) does not consider them compounds on the grounds that expressions with "coordinated elements are phrases" (in this case the coordinator would be implicit; for a broad view on coordinating constructions in typologically different languages, see Haspelmath 2004).

(36)   a. **actor-director**, author-illustrator, woman-doctor, fighter-bomber, he-cheetah, hero-martyr, jazz-rock, king-emperor, library-guestroom, maid-servant, man-servant, owner-occupier, panty-girdle, player-coach, player-manager, poet-translator, producer-director, scientist-explorer, screwdriver-hammer, secretary-treasurer, she-goat, singer-songwriter, sofa-bed, speaker-listener, washer-dryer, and worker-priest.

b. angel-beast division, black-white relations, the Cadbury-Schweppes business, the doctor-patient gap, father-daughter dance, grandmother-grandchild relationship, Harper-Collins, love-hate relationship, the love-pain equation, **the mind-body problem**, a modifier-head structure, the nature-nurture debate, north-west, parent-child relationship, Urbana-Champaign, and the Wellington-Auckland flight.

In B&S's view, the compounds in (36a) are endocentric. Semantically, they are said to be double-headed. There is a relation of coordination between the two nouns, both of which are understood as being hyperonyms of the compound: an *actor-director* is an actor and a director. Lieber (2008) calls this relationship 'simultaneous', and Olsen (2001, 2004) calls the compounds having this relationship 'copulative compounds'. Formally, the compounds seem to be right-headed structures: plural marking is placed on the second noun, as in *There are many poet-translators in this country* (example from Plag 2003: 147). However, if there is a CRD relation between the two nouns and both of them are interpreted as being hyperonyms of the compound as a whole, plurality must have scope over the two nouns. The conclusion must then be that despite the second noun being formally marked for plurality, the two nouns are formal heads. The

nominal status of such compounds can come from either element. The coincidence of semantic heads with formal heads thus derives the endocentricity of the compound.[138]

A different view is held by Levi (1978: 93-94), who believes that, despite the compounding nouns being in a coordinate relation, the resulting compound (or the 'complex nominal' in her terms) is exocentric because neither noun is the head semantically. She reaches this conclusion by proposing an underlying relative clause whose head is deleted. For example, she derives *speaker-listener* from 'person who is (both) a speaker and a listener', with *person* being deleted.

Despite apparent formal identity among the compounds listed in (36a), some can only be interpreted as ATR compounds. That is the case for the compounds whose first element denotes the gender of the noun in second position: *he-cheetah*, *maid-servant*, *man-servant* and *she-goat*. In these cases, speakers treat the compound as a hyponym of the second noun, which they regard as the semantic and formal head. The first noun, which basically has the function of the adjectives male and female (i.e. a sex-marker), is seen as an attribute. On this reading, the compounds would be endocentric ATR. A CRD relation may also seem odd for other compounds. For example, in the case of *worker-priest*, being a priest implies being a worker. That is, the word *worker* does not add any new information to the compound, and a CRD relation may seem odd. The same reasoning can be extended to *hero-martyr*: martyrs are assumed to be heroes. As for the remaining forms, although there is a tendency to treat them as CRD compounds (e.g. *actor-director*, *author-illustrator*, *poet-translator*, *producer-director*, *singer-songwriter*), not all of them are treated as such by native speakers. For example, some speakers treat *fighter-bomber*, *jazz-rock* and *player-manager* as endocentric single-headed compounds with the second noun acting as the semantic head (a hyperonym of the compound), which can also be taken as the formal head (plural marker, nominal category). The first noun acts as a modifier of the head noun giving it some properties. In short, it seems that the CRD relation is possible when the two compounding elements can equally contribute new information to the compound by their being semantically parallel. These requirements are not satisfied by compounds where the first element is a gender marker (*she-goat*) or contains information already present in the second noun (*hero-martyr*), but seem to be satisfied by compounds denoting two job titles (e.g.

---

[138] *Jazz-rock* cannot be pluralized, but the same treatment can be maintained (endocentric CRD compound). Both *jazz* and *rock* can be claimed to be formal heads by providing the syntactic category to the compound.

actor-director) or two types of devices/machines (*washer-dryer*) although not always (e.g. *fighter-bomber*). What these results suggest is that two apparently coordinated nouns can indeed be interpreted as coordinate but also as a modifier-modified structure, the final interpretation probably being subject to the speaker's knowledge of the world.

That said, we want to argue that when two nouns have a CRD reading, they do not form a compound (compare Adams 2001). We think that the symmetrical relation that is established between the two nouns is due to asyndetic coordination: an implicit conjunction is understood between the two nouns. Coordination is attested in syntax but it is not clear whether it exists in morphology. We assume that a true coordinate relation (e.g. an entity having properties of both A and B) can only be established in syntax.[139] Accordingly, NN forms with a coordinate relation will not be treated as compounds but as cases of syntactic coordination. However, as we have just seen above, NN compounds listed in (36a) like *fighter-bomber* and *player-manager* can be interpreted as endocentric single-headed compounds: the second noun is the head formally (plural marker, nominal category) and semantically (a hyperonym of the compound). As a result, the compound has a modification/subordination relation: the compound denotes a subset of the set of entities denoted by the head noun, which is given some properties by the first noun (the Catalan counterpart in (65c) is given the same treatment). As defined by native speakers, a *player coach* is 'a coach who is also a player on the team' or 'a coach that plays with the team' (ATR/SUB) and *jazz rock* is 'rock with some characteristics of jazz' (SUB). These compounds will be placed under endocentric SUB compounds provisionally (see the discussion section where the distinction between ATR/SUB is further elaborated upon).

Regarding the examples in (36b), they are exocentric CRD compounds in B&S's view. The two members of the compound characterize an entity outside the compound, with which they stand in a particular relationship, as in *the mind-body problem*, understood as the problem between the mind and the body. Lieber (2008) distinguishes three possible relations between the two constituents of the compound, which she calls relationship (*parent-child relationship*), collective (*father-daughter dance*) and disjunctive. The disjunctive relation cannot be exemplified with any of the examples here. In fact, the disjunctive relationship, which Lieber exemplifies with *pass-fail*, is not

---

[139] Our view is in agreement with authors like Bresnan & Mchombo (1995), who argue that coordination is syntactic (as opposed to morphological).

necessary. By appealing to pragmatics, we can observe that disjunction can be subsumed under relationship: one necessarily passes or fails. Similarly, Bauer (2008) also distinguishes different subtypes of compounds. More specifically, he distinguishes translative compounds (*the Wellington-Auckland flight*) from co-participant compounds (*parent-child relationship*). In the former, the order of the elements makes a difference in meaning since there is a starting point and a finishing point, and in the latter there is some interaction among the participants.

Contrasting with this view, we believe that the forms in (36b) are not exocentric CRD compounds but endocentric compounds with a subordination relation between the head and the non-head. As already discussed above for the forms in (36a), we treat as phrases NN forms with a coordinate relation. Such phrases cannot then form compounds by themselves but can be incorporated in the non-head position of a compound, as is the case of the compounds in (36b) (see pp. 104-109). Our proposal is that the forms in (36b) are compounds not by virtue of the CRD relation established between the elements constituting the phrase (as has generally been assumed) but by virtue of the subordination relation established between the phrase in the non-head position (which acts as a simplex word) and the noun in head position. To illustrate the point, in *mind-body problem*, *problem* is the head of the compound and *mind-body* is its non-head, which happens to be a syntactic phrase turned into a word and inserted in the non-head position of the compound. The specific relation between the elements of the compound will be determined by the semantics of the head (cf. e.g. Pustejovsky 1995): e.g. *the mind-body problem* refers to the problem of how the mind relates to the body, hence SUB compounds. (See Wisniewski 1996 for a different view of how these compounds get their interpretation, according to which basic concepts combine to form more complex ones).

If the compounds in (36b) were exocentric compounds, as B&S claim, they would be quite different from other compounds that are classified as exocentric in B&S's system, such as *butterfingers* (35a) and *redhead* (38b). These two compounds are said to be exocentric because their referent (the 'semantic head') is not determined by *fingers* and *head* (unlike my 'metonymy' analysis), but by an entity outside the compound, i.e. a type of person. However, B&S's explanation for exocentricity cannot be extended to any of the examples in (36b). For example, *mind-body* does not uniquely refer to a problem (only *mind-body problem* does). In my analysis, *mind-body* just means 'mind and/or/… body' and can be combined within an endocentric compound

with any noun to its right: *mind-body question* (referring to a type of question), *mind-body relationship* (referring to a type of relationship), *mind-body discussion* (referring to a type of discussion), *mind-body exhibition* (referring to a type of exhibition) and so on. The same can be said of the other phrases occupying the non-head position of the compound: *doctor-patient* does not uniquely refer to a gap and *father-daughter* to a dance (nor to a relationship, conversation, bond, conflict, etc.).[140] In contrast, it is impossible to combine a compound like *redhead* with a noun to its right that refers to the semantic head of *redhead* (e.g. *person*), since it would be semantically superfluous (i.e. the word 'person' is already implied)*: *redhead person*.

[VN]ₙ compounds

VN compounds are all SUB compounds, which can be divided into two different patterns: the compounds in (37a) are endocentric and those in (37b) have traditionally been considered exocentric, although the latter will be argued to be endocentric (see below). Both types are very restricted in productivity. When Spencer (2003b) refers to the compound types available in English, for instance, he does not mention the exocentric [VN]ₙ compound (37b) and he observes that the endocentric [VN]ₙ compound (37a) is exceptional. Similarly, when talking about the compounds in (37a), Lees (1960: 150-151) observes that "(…) the pattern hardly seems productive at present" (see Giegerich 2004: 3 for similar remarks). In addition, the majority of these compounds (37a-b) are lexicalized and not decomposable synchronically.

(37)   a. bakehouse, call girl, drophammer, glow-worm, launch window, pay day, playboy, playground, playtime, punch-line, rattlesnake, scatterbrain, scrubwoman, search engine, **search party**, swearword, think tank, tow-path watch-tower, and whetstone.

        b. catch-fly, cutpurse, cutthroat, daredevil, hangman, heal-all, killjoy, **pickpocket**, rotgut, scarecrow, spendthrift, spoilsport, tear-thumb, telltale, tumble-dung, and wagtail.

---

[140] One question that may arise from the previous discussion, though, is why a phrase, without an overt coordinator, is usually odd at best when used syntactically, but fine in the non-head position of a compound (*??mind-body is an interesting problem*). A tentative answer could be that a syntactic phrase must omit some material if it is to appear in the non-head position of a compound, as has been argued for telegraphic speech in newspaper headlines (see A&N 2004: 123, fn. 10 for similar discussion), whereas such material must be present in syntax.

The compounds in (37a) are endocentric: they identify a subset of the set denoted by the head noun (in second position) and formal markers attach to the head noun as well. For example, a *search party* is a party of people who search for someone. Being in nonhead position, the verb cannot have its argument structure satisfied, so the noun is not an argument of the verb and the function of the verb is to modify the head noun. As for the exact semantics of this compounding type, no common underlying pattern seems to exist. We can only predict a vague meaning of subordination since the subordination of the verb into the noun is different in each compound. Bauer (1983) notes that sometimes it is difficult to decide whether the first element is a N or a V, which may be the case for *checkpoint*, *showroom*, *wash-day*, and *dance hall* (examples from Adams 1973).

By contrast, the compounds in (37b) are generally claimed to be exocentric, a claim which is usually illustrated with paraphrases: a *pickpocket* is not a kind of pocket, but somebody who picks pockets. The compounds can denote people (e.g. *pickpocket*, *killjoy*, *spendthrift*), animals (e.g. *wagtail*, *tumble-dung*), plants (*catch-fly*, *tear-thumb*, *heal-all*) and objects (e.g. *rotgut*, *scarecrow*). The compounds which denote people have a pejorative connotation, which is absent in the compounds denoting animals, plants and objects. Recall the paraphrase of a *pickpocket*, which is a person who picks pockets. Unlike in Romance, this pattern is very limited in English: most compounds are lexicalised (Bauer 1983, Carstairs-McCarthy 2002, Plag 2003) and the new ones are limited to non-human denotata, like *Xpel-air*, which is a kind of fan (see Marchand 1969: 380-382 for more examples). Although there is a long tradition treating these compounds as exocentric, one could also argue that there is a zero-affix responsible for the nominal category and the semantics of the compound, thus deriving its endocentricity. This is the view that we defend and that will be further developed when discussing the Catalan counterpart (see subsection 2.3.2.1 for the [VN]$_N$ compounds in Catalan). Concerning the grammatical relation between the V and the N, and between the complex [VN] and the zero-affix, it is of a SUB nature. The noun is interpreted as the internal argument of the verb, and the [VN] is in turn subordinated to the zero-affix, hence the label of SUB compounds.

[AN]$_N$ compounds

The compounds with an [AN]$_N$ structure are all ATR: the adjective is attributed to the noun. Traditionally, a distinction has been made between endocentric (38a) and exocentric (38b) compounds. (For more examples of each type, see Lees 1960: 129-130).

(38)   a. avian sanctuary, blackbird, blackberry, blackboard, blackmail, bluejay, bovine disease, brownstone, classical music, dental appointment, dry cleaning, easy chair, fast-food, greenfly, greenhouse, greenstone, hard hat, hard-stuff, herbal remedy, High Court, hotbed, hothouse, narrow-boat, nervous system, new town, polar bear, poorhouse, quicksand, **red squirrel**, revolving door, sharpshooter, silly-season, smallpox, solar panel, sour-dough, tidal wave, tropical fish, urban transportation, wet-suit, White House, and wildfire.

   b. blackcap, bluebell, bluestocking, boldface, dimwit, fathead, greenback, **greybeard**, hardback, hard top, heavyweight, highbrow, lazybones, longlegs, longnose, loudmouth, paleface, redbreast, redcap, redcoat, redhead, redlegs, redneck, redshank, redskin, shorthorn, thick-head, wetback, whitebeard, whitethorn, and yellowtail.

The meaning of the compounds in (38a) is not fully compositional: *easy chair* is a kind of chair (semantic head), but what kind of chair it is cannot be predicted from the sum of the meanings of the two elements. Plural marking is placed on the noun, which determines the categorial status of the compound (formal head). According to Lieber (1983: 255, 260), AN compounds of the type in (38a) are productive, but we believe that this type is not as productive as NN compounds (e.g. *bell jar* (35b)) and in fact its productivity is quite restricted. There are a few lexicalised compounds of this type and the range of adjectives that can occur in first position is limited, mainly to monosyllabic adjectives of Germanic origin. If some of them do not conform to this restriction, they may be early Romance loans like *double talk* (Bauer 1983). In addition, we find it difficult to draw the line between AN compounds and phrases. In a closely-related language like Dutch, by contrast, it is easier to distinguish them: the adjective is inflected in the phrase but uninflected in the compound. One could try to find out the

status of some English AN forms by looking into the status of their Dutch counterparts, a task not undertaken here, though.

Adams (2001: 81) does not consider AN sequences with a gradable adjective (e.g. *complex, long*) as compounds: e.g. *a still outstanding claim*, in which the A is modified. She only considers compounds those AN sequences in which the A is not gradable (e.g. *military, rural, herbal, editorial*) and has the same function as modifying nouns: compare *military sales* vs. *arms sales*, *country pursuits* vs. *rural pursuits*. Note, however, that AN sequences with a relational adjective (or an 'associative attributive' in Huddleston & Pullum's 2002 terms) like *herbal* and *polar* cannot be considered compounds but phrases (see e.g. Giegerich 2004: 13, 2005: 587 for more examples). The possibility of applying the *one*-replacement test shows their syntactic nature. That is, compounds do not allow their elements to be picked up anaphorically, but the forms containing relational adjectives in first position do. For instance, *panels* can be picked up anaphorically, as in *We are not using solar panels but lunar ones*. In short, after removing these syntactic forms, not many AN compounds of the type in (38a) are left, which makes us question their existence. They are included within parentheses in Table 2.1.

Traditionally, the alleged exocentricity of the compounds in (38b) is explained by saying that the semantic head lies outside the compound: a *greybeard* does not denote a kind of beard but a kind of person who has a grey beard. However, we will maintain that the compounds in (38b) are endocentric and argue that their apparently exocentric interpretation arises from metonymic processes which lie outside the morphological component, in the same way as ATR compounds of the [NN]$_N$ type, like *butterfingers* (35a) (see also Olsen 2001: 312, fn. 3, and the references therein). The noun is the head both formally and semantically: it is the bearer of plural marking and the adjective gives an attribute to the noun (i.e. the beard is grey in *greybeard*). Note that there are compounds like *lazybones* which are plural on the surface but can refer to either one or more than one person. In these cases, the noun is inherently plural and makes it impossible to know when the compound is being used in singular or plural because when pluralized, the plural marker of the compound and the plural marker of the noun are realized on the same head and are fused into one *–s*.

According to Carstairs-McCarthy (2002) and Giegerich (2004), this type of compound is not productive, and according to Plag (2003), the compounds can refer to

human beings (*dimwit, greybeard, lazybones, paleface*) or higher animals (*longlegs, longnose, redbreast, shorthorn*), but as the examples show they can also refer to plants (*bluebell, whitethorn*) and objects (*greenback, hardback*), although to a smaller extent. Note that the compounds which refer to people have a negative connotation, in the same way as the compounds of the *pickpocket* (37b) type. (See Marchand 1969: 386-389 for more examples, some of which are old-fashioned nowadays).

[PN]_N compounds

[PN]_N compounds are difficult to accommodate into B&S's classification. The following treatment (especially the ATR/SUB distinction) should be taken as provisional (to be further commented upon in the discussion section). [PN]_N compounds seem to be divided into ATR and SUB, ATR compounds being all endocentric (39a). For example, as defined by a native speaker, an *under-pass* is 'a road underneath a bridge, or a tunnel through a mountain', and an *outpost* is 'a station (e.g. military or exploratory) remote from the main quarters', *out* indicating the remoteness of the post. Plural marking is placed on the noun. As for SUB compounds, some are exocentric (39b) and some are endocentric (39c). In the case of exocentric compounds (39b), the compound refers to an entity outside the compound which is characterized by the compound: an *underground* refers to a railway system that is under the ground (typically) and an *underarm* refers to the area under one's arms, i.e. to the armpit. For such compounds, plural marking should be understood on the exocentric head. Concerning endocentric SUB compounds (39c), the noun is the head of the compound both formally (i.e. the noun inflects for plurality) and semantically, since the P is subordinated to the head noun: as defined by native speakers, an *in-joke* is 'a joke only understood by a select few people who are *in the know*', and an *out-tray* is 'a tray whose contents are ready to go out of the office'.

(39)   a. aftereffect, afterlife, afterthought, aftertaste, down-pipe, incrowd, ingroup, off-islander, outbuilding, **outpost**, outroom, overcoat, through-road, underbrush, undercoat, underhair, under-pass, and uptrend.

       b. afterbirth, **underarm**, underbelly, undergraduate, and underground.

       c. in-joke, and **out-tray**.

Some compounds may seem misplaced: one might argue that compounds like *aftereffect*, *afterlife*, and *afterthought* (39a) could be interpreted as exocentric SUB compounds like the compounds in (39b), and consequently an *afterthought* would be 'a thought that comes after the (first) thought', but as the paraphrase hints at, such a possibility is excluded. While it seems possible to have an exocentric head identical in shape to the one present in the compound (e.g. an *afterthought* is a type of thought), it seems less plausible that the SUB relation inside the compound can be implemented with material not present in the compound: e.g. *first* in the paraphrase of *afterthought*, 'a thought that comes after the (first) thought' but crucially what follows *after* does not need to be a thought, as a native speaker's paraphrase of *afterthought* reveals: 'a thought that occurs to you after you have made a decision/statement'. The upshot is that such compounds are best analyzed as endocentric ATR: the compound refers to a type of entity denoted by the noun, which inflects for plural marking, with the P giving an attribute to the noun. For instance, an *afterlife* is a life after the present life, and an *aftereffect* is an effect that occurs after an event.

In fact, we want to claim that there are no exocentric compounds and treat the compounds in (39b) as endocentric. Although it is not obvious at first sight, we want to claim that the head on which the plural marker is realized is also responsible for the nominal category of the compound (formal head) and its semantics (semantic head) (see the Catalan counterpart in (69a) which receives a similar analysis). Such treatment will give uniformity to the compounding process. Note that the relation between the two visible elements can be compared to the relation established in P+N forms like *outdoors*, *offstage* and *uphill* presented in subsection 2.3.1.

Some cases of PN compounds, such as *oversight*, *underdog* and *uprising*, have become lexicalized and the compound is no longer treated as the union of a P and a N: e.g. an *uprising* is understood as a rebellion. Also note that some prepositions can be more easily combined than others with nouns to form compounds (e.g. *after*, *in*, *out*, *over*, *under*) and that cases very similar to the ones presented above have been excluded since they are cases of prefixation rather than compounding: prefixed nouns are more evident when the noun can be related to a verb, like *overdose*, *overkill* and *overtax*. In such examples *over-* has developed a meaning different from the meaning of its independent counterpart. The bound form seems to quantify over the event implicit in the noun: *overdose*, *overkill* and *overtax* express the underlying verb (*dose*, *kill*, *tax*) in

an excessive quantity. In this respect, a speaker's paraphrase of *overdose* is revealing: 'too much of a medicine/drug, exceeding the recommended dosage'. Similar paraphrases were given for *overkill* and *overtax*. That this hypothesis is on the right track seems to be confirmed by the findings in Berg (1998), which show that all cases of [PN]$_N$ with a derived deverbal noun come from [PV]$_V$ originally, which were later converted to nouns. Other examples of prefixation where *over-* and *under-* attach to deverbal nouns are *overcompensation, overplanning, over-expansion, underconsumption, underfulfilment, underfunding*, and *under-ventilation* (cf. Adams 2001: 75-76). These examples further confirm our hypothesis that the prefixes *under-* and *over-* act as prefixes: they quantify over the verb underlying the noun and the *over-/under-*prefixed verbs can be paraphrased as 'to V in an excessive/insufficient way'. Similarly, forms like *off-cut* and *out-take*, superficially similar to the compounds listed in (39), are also excluded: they are nominalized phrasal verbs, not cases of PN compounds.

In short, not all cases of [PN]$_N$ combinations are compounds. They can be prefixed words when the P has developed a meaning different from the meaning of the P when it is found in isolation, and the P can form a series with the same meaning, as we have seen for *over-* and *under-*. [PN]$_N$ combinations can also be phrasal verbs which have undergone nominalization and whose constituents have been inverted. PN compounds have been divided into SUB and ATR endocentric (but see the final discussion of such a division in section 2.4).

[NA]$_N$ compounds

The compounds in this group are all endocentric ATR compounds, with the head on the left, an unexpected fact for English given the RHR. The left-headedness of such compounds is explained by the fact that they are based on Romance compounding, where this pattern is attested. They are not productive in English, whose counterpart would have the opposite order of constituents in the compound and the head would be on the right, namely they would be endocentric ATR [AN]$_N$ compounds (e.g. *red squirrel* (38a)). Due to the weight of the RHR in English compounding, sometimes there is variation regarding the placement of plural markers: *attorneys general* vs. *attorney generals*.

(40)   **attorney general**, heir apparent, notary public, and solicitor general.

[VV]$_N$ or [V[V]$_N$]$_N$ formations

Nominal complex forms of the [VV]$_N$ or [V[V]$_N$]$_N$ type are almost non-existent and they can also have a verbal or adjectival use, in addition to the nominal one. We will maintain that they do not constitute a compound type. Consider the forms in (41).

(41) make-believe, shrink-wrap, slam-dunk, and strip-search.

Make-believe, which can also be used as an adjective and verb, seems to be a nominalization of a verbal syntactic phrase (see verbal compounds of the VV type). An example of *make-believe* used as a noun is as follows: "A fiction writer's childish willingness to immerse himself in *make–believe* — John Updike".[141] Speakers prefer *shrink-wrap* as a verb, but those who accept it as a noun interpret it as a plastic film for wrapping stuff, with *shrink* not playing any role. *Wrap* gets the plural marker when the form is pluralized. When *slam-dunk* and *strip-search* are used as nouns, the interpretation is that of an endocentric SUB compound. *Dunk* and *search*, verbs treated as nouns, get the plural marker, and *slam* and *strip* are understood as actions subordinated to the noun. For example, a *strip-search* is a search of a person who is made to undress. *Slam-dunk* and *strip-search* can be seen as conversions of verbal compounds. (For other examples like *drop-kick* and *stir-fry*, see subsection 2.3.1.2 where [VV]$_V$ compounds are discussed). In short, it seems that there is no general compounding process of the [VV]$_N$ or [V[V]$_N$]$_N$ type.

[VP]$_N$ formations

Most complex words in (42) can be related to phrasal verbs (e.g. *His marriage **broke down** soon after they had a child*), but there is no one-to-one correspondence: *pray-in* and *teach-in* do not come from phrasal verbs.

(42)   breakdown, call-up, drawback, fallout, kick-off, lie-in, laugh-in, love-in, make-up, play-back, press-down, put-down, put-on, read-through, runaway, sell-out, sit-in, sleep-in, pushover, stopover, take-over, take-off, talk-in, think-in, warm-up, washout, wrap-up, and write-off.

---

[141] Example from *Merriam-Webster Online Dictionary*: MWOD.

There seems to be a division between those [VP]$_N$ formations which come from phrasal verbs which subsequently undergo conversion to nouns (the vast majority of cases) and those [VP]$_N$ formations which do not come from phrasal verbs but arise from the union of a verb and a particle such as *V+in*, which became fashionable in the 1960s and could be considered a case of suffixation. Initially *V+in* formations denoted 'group protest', which was later replaced by a connotation of 'group activity'. Other particles that seem to function in the same way as *V+in*, in the sense of being part of formations in a series with a specialised meaning, are *blackout*, *brown-out*, *dim-out* and *white-out* (cf. Adams 2001: 77).

Berg (1998) agrees with the general view that behind forms like *breakdown$_N$* there is no regular compounding process, but a process of conversion from a phrasal verb into a N, which is evidenced by a change of stress: compare [*to break dówn*]$_V$ with [*a bréakdown*]$_N$. While acknowledging that forms like *breakdown$_N$* are the result of conversion, authors like Carstairs-McCarthy (2002) argue that they are real compounds although "marginally productive" while other authors like Bauer (1983) do not consider them compounds in the strict sense.

The position taken in this work is that [VP]$_N$ formations are not compounds. Those that come from phrasal verbs are created in the syntax, i.e. they are a syntactic product which is later converted into a noun, and those that involve a particular P which develops a specific meaning can be considered the product of a special kind of affixation. Evidence for the latter type has already been given (the formations in a series) and further evidence for the former type (42) is provided by Adams (2001: 76). She provides sentences like (43) where the nominalized phrasal verb is being used as non-count and (44) where "they denote an instance of the verb's action following *have/give/take a*":

(43)   a. We were supposed to keep out of the pilot's way at blast-off (1952).

         b. Gas will be liberated… during pumpdown (OED: 1971).

(44)   a. give (something) a rub-down

         b. have a fry-up, a punch-up, a sleep-in

[PV]_N formations

The same explanation given for [VP]_N formations coming from phrasal verbs also applies to [PV]_N formations, like the ones given in (45):

(45)   downfall, intake, and upkeep.

These forms also come from converted syntactic constructions: [PV]_V or [VP]_V, the latter involving inversion of constituents in addition to conversion to a noun. In the case where there is no verbal syntactic counterpart for some of the [PV]_N formations nowadays, Berg (1998) claims there was one in the past. In conclusion, no regular compounding process seems to exist for [PV]_N complex words.

This concludes the survey of nominal compounds in English. The results so far are gathered in the following table (Table 2.1): one example of each type is given and where we distinguish several subtypes, one example of each is also provided. For example, when discussing the [NN]_N type, three subtypes were mentioned (N2 being a relational noun, a deverbal noun or neither of the two previous cases) and hence they are all exemplified. The [VV]_N/[V[V]_N]_N, [VP]_N and [PV]_N forms have not been included, because they have been argued not to be compounds, and the examples of the compound types with limited productivity are included within parentheses.

Table 2.1: Nominal Compounds in English

	NOMINAL COMPOUNDS					
	SUBORDINATE		ATTRIBUTIVE		COORDINATE	
	endocentric	exocentric	endocentric	exocentric	endocentric	exocentric
[NN]$_N$	table leg (32a) bookseller (32b) clog dance (32c) actor-director (36a) mind-body problem (36b)		(butterfingers) (35a) bell jar (35b)			
[VN]$_N$	(search party) (37a) (pickpocket) (37b)					
[AN]$_N$	F.-G. cooperation (55b)		(red squirrel) (38a) (grey beard) (38b)			
[PN]$_N$	(underarm) (39b) (out-tray) (39c)		outpost (39a)			
[NA]$_N$			(attorney general) (40)			

2.3.1.2 Verbal compounds

Marchand (1969: 100-107) calls verbal compounds which are formed by a noun or an adjective and a verb pseudo-compounds. In the literature, it has often been claimed that verbal compounds in general are rare and those existing are mostly derived via conversion or backformation (e.g. Adams 1973, 2001, Booij 2005, Plag 2003, a.o.). On a similar note, Spencer (2003b) observes that there are only a few exceptional examples of base-generated verbal compounds, such as *sight sing* and *sight read* of the [NV]$_V$ type and *drink-drive* of the [VV]$_V$ type. However, Marchand remarks that verbal compounds may be common in specialized jargon. He mentions that, for example, words like *stallfeed*, *smokedry*, *winterfeed*, and *winterkill* belong to the jargon of farmers. Also, against the general assumption, Bauer & Renouf (2001) observe that, in addition to the [PV]$_V$ type (the only type acknowledged by Selkirk 1982), there are other types, which are present in their 1988-1998 corpus of the British newspaper *The Independent*. They mention, as cases of verbal compounds, *outsoap* and *out-Herod*, of the [PN]$_V$ type (*outsoap* could also be [PV]$_V$), the form being more productive when the N is a proper name. While these may be argued to be cases of prefixation (as we will indeed argue), Bauer & Renouf provide other compounds which conform to the verbal compounding pattern, namely [N/V/A+V]$_V$: *custom-produce*, *thumb-strum* (NV), *dry-burn*, *freeze-dry* (VV), *slow-bake*, *hardwire* (AV).

As will shortly be seen, our position is that despite the fact that some verbal compounds may be the result of backformation, their status in the grammar cannot be any different from cases that arise spontaneously. First, it is unlikely that the process of backformation can result in an acceptable object, namely a verbal compound, if the principles of English morphology do not permit such a type of compound. Second, there are quite a few verbal compounds which are recent coinages. Regarding NV compounds, Bauer (1983: 208) notes that "There are plenty of this type of verb being coined in current English, some recent examples being *blockbust, carbon-date, colour-code, head-hunt, sky-dive*". Since a nominalizing suffix can be added to any verbal compound, one could argue that verbal compounds are all derived from the nominal forms by means of backformation. Although unfalsifiable, this position seems to be an easy way out. Finally, there is no way for children learning English to know whether a verbal compound is base-generated or is the result of a backformation unless they are given explicit evidence, which is unlikely, to say the least (cf. subsection 1.4.2.2 in chapter 1; see also McIntyre 2009 for the same results).

This subsection contains three different types of verbal compounds, namely [NV]$_V$, [VV]$_V$, and [AV]$_V$ compounds, which are followed by two different types of constructions, i.e. [PV]$_V$ and [PN]$_V$ complex forms, which are taken by some authors to be compounds but we conclude that they are better analysed as a different construction. A table summarizing the results ends this subsection.

[NV]$_V$ compounds

As already noted in the introduction to verbal compounds, [NV]$_V$ compounds are often considered exceptional under the general assumption that there is no general process of [NV]$_V$ compounding in English (which may explain why Selkirk 1982 denies their existence) and that the attested cases are mostly attributed to backformations from nominal or adjectival compounds (e.g. *proof-reading* or *proof-reader > proof-read*, *talent-spotter > talent-spot*, and *machine-washable > machine-wash*) or the result of a conversion process (*litmus-test*$_N$ → $_V$, *handcuff*$_N$ → $_V$,   *mountain-bike*$_N$ → $_V$) (cf. Adams 2001, Bauer 1983, Bloomfield 1933, Booij 2005, Carstairs-McCarthy 2002, Plag 2003). Despite the fact that some [NV]$_V$ compounds may be derived, we maintain that they have the same status as those that are base-generated, although it is difficult to know which is which since all verbal compounds have a nominalized counterpart. Some examples of base-generated forms might be *chain-drink, chain-smoke, sight-read, sight-sing, sight-translate* (Spencer 2003b). All verbal NV forms will be equally treated as compounds. Note, though, that the use of a finite form in a sentence is worse than a non-finite form like a gerund: *Mountain-climbing is good for one's health* is better than *I mountain-climb every weekend*.

Verbal compounds of the [NV]$_V$ type are divided into endocentric SUB and endocentric ATR compounds, exemplified by (46a) and (46b) respectively. The verb is the head both formally and semantically: it determines the verbal status of the compound and formal marking is placed on it (e.g. the verb inflects for subject-verb agreement, tense). Semantically, the compound denotes a subtype of action denoted by the verb. The endocentricity of the compound is thus derived.

(46)   a. air-condition, babysit, base-generate, book-keep, brainwash, breast-feed, browbeat, carbon-date, cheer-lead, colour-code, **computer-generate**, custom-produce, earmark, gift-wrap, globe-trot, handcuff, hand-make, hand-wash, hand-weave, head-hunt, housekeep, litmus-test, machine-wash, mass-produce, moon-

light, mountain-bike, parcel-bomb, pressure-clean, proof-read, rugby-tackle, sightsee, sight-read, sight-translate, sky-dive, spoon-feed, spot-light, spring-clean, stage-manage, steam-clean, talent-spot, tape-record, thumb-strum, volume-expand, and window-shop.

b. chain-drink, **chain-smoke**, and ghost-write.

Regarding the compounds in (46a), the N inside the compound is not argumental, which is clear from examples like *computer-generate* and *steam-clean*, where the N can be understood as the instrument with which one performs the action (e.g. to clean by means of steaming) (47a). When it looks like the noun performs the function of the internal object of the verb (e.g. *brain-wash*, *talent-spot*), the compound verb can take an external object in syntax (e.g. *They babysat John all afternoon*), which means that the noun inside the compound cannot be taken as the internal argument (47b).[142] This is in utter contrast with NV compounding in Frisian, where the noun is argumental in complex verbs like [[*messe*]ₙ [*slypje*]ᵥ]ᵥ (knife+sharpen), since the simple transitive verb becomes intransitive when complex and no external object is allowed (for details, see Dijk 1997). NV compounding in English can also be contrasted with NV compounding in Dutch where the N can be argumental when the verb strands the noun under V2 (e.g. *koffie+zetten* (coffe+set) 'make coffee') or non-argumental when the complex verb moves as a whole under V2 (e.g. *slaap+wandelen* (sleep+walk)) (on this point, see Ackema 1999b).

(47)   a. In order to computer-generate logic diagrams corresponding to a text, the logical structure must be evident to the device performing the task. (…)

---

[142] We have found only two examples which seem to contradict such a statement. The verb *to housekeep* is one of them. The MWOD defines *housekeep* as 'to perform the routine duties (as cooking and cleaning) of managing a house' and analyses it as an intransitive verb. The verb *keep* is transitive and when it is compounded with *house* becomes intransitive, which seems to indicate that *house* functions as the internal argument of the verb (e.g. *The old lady housekept when we were children*). The other contradicting example is *to walk organize*, which was used by a Scots native speaker (in 2008) in the following sentence:

(i)   Also if you did not try to get on the trip but could be tempted to walk organise let me know and I will get back to you about whether you have got on the trip or not!

In the absence of more data (despite these two contradictory examples), we can maintain that the general statement made above holds: the N in NV compounds is not argumental in English. Whether this is a possibility in Scots, for example, is left for future work.

b. Besides training young cyclists, organisers are also trying to talent-spot members to form a national team for the YOG.

As for the endocentric ATR compounds in (46b), the nouns are mostly understood as involving a comparison. For example, *chain-smoke* means to smoke cigarettes one after another like a chain. An example of such a type of compound follows:

(48)   Sir Sean had most recently commissioned Hunter Davies, the only authorised biographer of the Beatles, to ghost write the book.[143]

The productivity of such compounds is more limited than the compounds illustrated in (46a). There are very few examples and speakers find it difficult to make up new forms: e.g. *He **king-marched/king-walked** down the street*.

[VV]$_V$ compounds

Verbal compounds of the [VV]$_V$ type are regarded as non-existent by Selkirk (1982), as exceptional by Bauer (1983) and Spencer (2003b), or are simply not mentioned, as in Katamba & Stonham (2006). For example, Bauer (1983: 208) argues that the attested examples (e.g. *typewrite, test-market*) are dubious in that they do not inevitably belong to this group. Like [NV]$_V$ compounds, [VV]$_V$ compounds may be related to nominal or adjectival compounds by means of a suffix (e.g. *crash-landing*$_N$ ~ *crash-land*$_V$, *dive-bomber*$_N$ ~ *dive-bomb*$_V$, *dry-cleanable*$_A$ ~ *dry-clean*$_V$, *sleep-walker*$_N$ ~ *sleep-walk*$_V$) or without (e.g. *drop-kick*$_{N/V}$, *slam-dunk*$_{N/V}$, *shrink-wrap*$_{N/V}$, *strip-search*$_{N/V}$), which may lead one to think that there is no [VV]$_V$ compounding process. Even though one may think that some compounds (e.g. *freeze-dry*, *drink-drive*, *dry-burn* and *fly-drive*) can be regarded as base-generated VV compounds (see Lieber 1983: 265 for other examples), they can all have a nominal or adjectival counterpart as well. As noted above, it is a fact about English morphology that verbs can be related to nominal and adjectival forms by means of suffixes (e.g. the adjectivalizing suffix *-able* and the nominalizing suffix *–ing*). From this perspective, one could argue that verbal compounds are always the result of backformations (i.e. of derived adjectival and nominal forms). Despite being an unfalsifiable claim, we believe that the grammar is unlikely to result in an acceptable

---

[143] Examples (47a, b) and (48) are the result of a Google search.

object, namely a [VV]$_V$ compound, if the grammatical principles do not allow such a type of object (cf. the introduction to verbal compounds in the present subsection). In short, despite the small number of attested forms, we will take VV compounds as a compound type available in English (although not as profitable[144] as endocentric SUB NN compounding like *clog dance* (32c)) (cf. Booij 2005, Plag 2003). Some examples follow:

(49) a. crash-land, dive-bomb, drink-drive, drop-kick, dry-burn, fly-drive, **freeze-dry**, shrink-wrap, slam-dunk, sleep-walk, stir-fry, and strip-search.

  b. trickle-irrigate and type-write.

  c. daresay, make do, and make-believe.

At first sight the forms in (49a) could be classified as endocentric CRD compounds in B&S's (2005) classification. Concerning the CRD relation, the conjunction 'and' would mediate the relation between the two verbs, of which the action of the first verb is understood as taking place first and that of the second verb coming after for most speakers. For example, *dive-bomb* involves diving first and bombing second, and in *strip-search*, first someone strips, and then you search them (a simultaneous relation in Lieber's 2008 terms). One speaker, though, regarded either order possible for *stir-fry* and *strip-search*, and another speaker gave a simultaneous interpretation to *stir-fry* and *crash-land*. However, most speakers understand the complex words as denoting a subtype of the type of action denoted by the second verb (semantic head): e.g. *crash-land* expresses a type of landing and *dive-bomb* a type of bombing. Formally, either verb could be responsible for the verbal status of the compound but note that speakers only inflect the second verb for past tense,[145] which indicates that the second verb is taken as the formal head. Where there is some indication of the past tense of such compound verbs in dictionaries (e.g. the *Datasegment Online Dictionary*: DOD), the second verb also seems to be the formal head: *freeze-dried*, *shrink-wrapped*, *strip-searched*, which agrees with the speakers' judgments. A search on the internet gives

---

[144] Bauer (2003) uses the term 'profitable' to indicate how much a process is used in the language.
[145] Speakers do no agree on the past tense form of *drink-drive*: some would not use *drink-drive* in the past, and others would inflect both verbs: *drank-drove*.

mixed results[146]: for example, it gives 60,900 hits for *sleep-walked*, 11,000 for *slept-walked* and 2,440 for *slept-walk*. Although the internet search indicates that past tense marking is possible on the second verb, on the first one and on both, there is a clear preference for attaching the past tense marking on *walk*, which may indicate that speakers take *walk* as the main verb semantically and formally. The result is that such compounds are endocentric: the second verb is both the semantic and formal head.

As already noted for the expressions in (36a, b), we understand complex forms with a coordinate relation not as compounds, but as phrases. If the forms in (49a) involve a true coordinate relation, they cannot be included in the study of English compounding. Although the presence of asyndetic coordination is a real possibility for some forms (e.g. *stir-fry*), speakers' interpretations show that this is not the only reading available. The forms in (49a) can also be analysed as compounds with the second verb being the head formally (i.e. inflection attaches to it) and semantically (i.e. the compound denotes a kind of action expressed by the second verb), and the first verb being a kind of manner/temporal modifier. Accordingly, *to dive-bomb* is expected to mean 'to bomb in a diving fashion/when diving', that is, a type of bombing. This expectation agrees with the interpretation given by speakers (see above). The fact that speakers interpret the action of diving coming first and the action of bombing second follows from the compounds being endocentric SUB: the action of the first verb is subordinated to the second verb but must be carried out first in order for the second verb to take place (e.g. in order to be able to bomb, one must first dive), which is also the case for the rest of the compounds.

The forms *type-write* and *trickle-irrigate* in (49b) also seem to be endocentric SUB compounds like the compounds in (49a). In *type-write*, *write* appears to be the semantic and formal head. That is, *type-write* can be interpreted as writing by typing and *write* is inflected for past tense. However, most speakers treat *type-write* as writing using a typewriter, the form from which the verb is derived. In other words, *type-write* is treated as a backformation. However, recall from the discussion above that forms like *type-write* can be assimilated to the group of endocentric SUB compounds with a VV structure (49a). Concerning *trickle-irrigate*, it also looks like an endocentric SUB compound: *trickle-irrigate* would be to irrigate in a certain way, namely by trickling.

---

[146] Searches on Google give rough results and should not be taken as definitive, but can certainly be indicative. The search on *sleep-walk* was carried out on 30th July 2008.

On closer inspection, *trickle* can also be treated as a noun ('to irrigate using only a trickle of water') and *trickle-irrigate* as a backformation of *trickle-irrigation*. In short, *trickle-irrigate* could be assimilated to the compounds in (49a) or (46a), given that the two patterns exist independently of backformations.

As for the forms in (49c), they are rather peculiar. Whether they are compounds or not is not obvious. If they can be treated as compounds, the alleged compounding process to which they belong is no longer alive in the language. They would be SUB compounds: the first verb selects the second one in all three cases. It seems that *daresay* should be distinguished from *dare-say*. *Daresay* can only be used in 1st person singular and in present tense (prescription found in the MWOD and observed by speakers), which suggests that it has become a lexicalised expression. Note that as a lexicalized form, if it could occur in other contexts other than present tense, we would expect formal endings to attach at the end of the word. This hypothesis is supported by a Google search, which gives 1,610 hits for *daresaid*, 1 for *daredsaid* and 22 for *daredsay*. This contrasts with the results found for *dare-say*: 16,500 hits for *dare-said*, 1,040 for *dared-said* and 40,200 for *dared-say*. *Dare-say* seems to be better analysed as *dare* being the head that selects a verbal complement, namely *say*. If this is the correct view, we would expect formal endings to attach to *dare* and not to *say* (a kind of 'want + to + infinitival form', or 'help + infinitival form') This expectation is confirmed by the results given above, which have the highest number of hits for *dared-say*, but the other possibilities of formal marking are also surprisingly quite high. The 16,500 hits for *dare-said* gives a blurred picture of the reality. Most of the cases consist of *Dare* as a proper name followed by the past tense of *say*. The Google search draws no distinction between the two forms being hyphenated or written as two separate words, the latter case also involving *Dare* as a proper name most of the time.

Make-believe and *make-do* are similar to *dare-say* in the sense that the first verb seems to select the second one. If they were compounds, they would be endocentric SUB compounds as well. *Make* gets the formal ending when the compound form is conjugated in the past in both forms. Consider the following examples, which are both from DOD:

(50)   a. He made believe that he didn't hear her

       b. They made do on half a loaf of bread every day

Rather than compounds, though, they resemble the causative construction in Romance: the first verb selects the second one in a sequence of two verbs. *Fer creure* would be the Catalan counterpart of *make-believe*, and *fer allargar/fer arribar* (make lengthen/make arrive) would correspond to *make do*. If this view is correct, the forms in (49c) should be dealt with in syntax and would not be compounds.

In short, verbal VV compounds are endocentric SUB compounds. Some sequences of two verbs that resemble verbal VV compounds in some respects have argued not to be compounds.

[AV]$_V$ compounds

Like verbal NV compounds, verbal AV compounds are usually claimed to be non-genuine compounds by the fact that some of them may be derived from nominal or adjectival compounds, via back-formation (e.g. *literary-editor > literary-edit*, *free association > free associate*, *soft landing > soft land*) or conversion (*blackmail, cold call, cold shoulder, free fall, short-circuit, shortcut, wisecrack*) (cf. Bauer 1983, Plag 2003). There are other cases where the verb and the noun arise almost simultaneously in the language, which makes it difficult to tell which compounding type is the genuine one. According to the MWOD, that is the case of *blue-pencil* (first attested as a noun in 1886, and as a verb in 1888), and *deep-freeze* (first attested as a verb in 1943, and as a noun in 1948), for example. All these facts put together may explain why verbal AV compounds have been called 'pseudo-compound verbs' by Marchand (1969) or claimed to be non-existent by Selkirk (1982). According to Plag (2003: 154-155), the main constraint for this type of compound comes from the fact that English verbs cannot have 'adjectival/adverbial non-heads', which he illustrates with the contrast between **fastdrive* and *a fast-driving chauffeur*, and **slow(ly)-move* and *a slow-moving animal*. Like in the previous two cases of verbal compounds (i.e. NV and VV), it might seem that there are cases in which there is no nominal/adjectival counterpart (e.g. *deep-fry*) or if there is one, it seems to be derived from the verbal compound (e.g. *double-book$_V$ > double booking$_N$*, *fine-tune$_V$ > fine-tuning$_N$*, *white-wash$_V$ → $_N$*), which could be taken as evidence for a genuine AV compounding process. The reality, nonetheless, points to the opposite direction: all AV compounds seem to have a nominal/adjectival counterpart (*deep-fry ~ deep-frying*). However, this fact cannot mean that AV compounds are not a legal object in the language; otherwise, nominal and adjectival AV forms would not

result in a verbal compound of the AV type (recall the discussion above). Therefore, all AV compounds (base-generated and derived) receive the same treatment in the present thesis, some of which are given in (51).

(51)   blackmail, blindfold, blue-pencil, broadcast, cold-call, cold rinse, cold-shoulder, deep-freeze, **deep-fry**, double-book, dry-clean[147], fine-tune, free-associate, free-fall, literary-edit, sharp-shoot, short-circuit, shortcut, short spin, slow-bake, slow-cook, **soft-land**, sweet-talk, quick-brew, warm iron, whitewash, and wisecrack.

Among the examples, note that forms like *cold rinse*, *short spin*, *slow bake* and *warm iron* seem to be derived from commands and the adjective seems to perform the function of an adverb. Recall from chapter 1 (subsection 1.4.2.2) that we follow authors like Emonds (1976) in that we treat adverbs as a derived category, the base form being an adjective. Although we understand that the term 'adjective' includes the categories which are traditionally defined as 'adjective' and 'adverb', at some points it will be useful to keep the two terms separate (as will be seen below, when revising Plag's 2003 constraint for AV compounding, for example).

All things considered, it seems that this compounding process is very limited and that, following B&S's classification, two types of compounds can be distinguished. On the one hand, some AV compounds are endocentric SUB compounds: *deep-fry* is 'to cook/fry in deep fat' (SUB relation and semantic head) and tense marking is placed on the verb (formal head). On the other hand, other AV compounds seem to be endocentric ATR compounds: *soft land* is 'to have a soft landing' (ATR relation: the landing is soft). In this case, *land* is also interpreted as the semantic and formal head: *soft land* is to land in a certain way, and the verb gets the verbal inflections. The difficulty in distinguishing between ATR and SUB compounds will be discussed in section 2.4, when the tripartite classification of B&S (2005) will be reconsidered and the conclusion will be that the ATR/SUB division is a fictitious one (although for the moment, the distinction will be maintained).

---

[147] We reject the treatment of *dry-clean* as a verbal V+V compound, an analysis available in the literature. Instead, we treat it as a verbal A+V compound, which is in agreement with speakers' judgments: 'to clean something in a way that keeps it dry' (compare Carstairs-McCarthy 2002).

Taking up the constraint mentioned by Plag, we believe that his statement that English verbs cannot have 'adjectival/adverbial non-heads' follows from A&N's (2004) morphosyntactic competition: if an adverb occupies the non-head position of the verbal compound [Adv-V], and the semantics of the compound is transparent, then competition between morphology and syntax predicts that the syntactic merger of the adverb and verb will win over the morphological one. The prediction is borne out by the data: *to deeply-fry* vs. *to fry deeply*. By contrast, if the compound contains an adjective and the phrase an adverb, then there is no competition, and the two structures are predicted to co-exist: *to quick-brew* and *to brew quickly*. Notice that sometimes a verbal AV compound (e.g. *He quick-brewed the stout*) is inserted in a larger compound: *quick-brew wine*, *low-build tyres* and *slow-cook marathon*.

Now we want to revisit Plag's claim that the contrast between *slow(ly)-move* and *a slow-moving animal* is due to the verbal and adjectival nature of *move* vs. *moving*. We agree that *move* is verbal and *moving* adjectival, but we think that the contrast in grammaticality has to be established on the basis of different sets of data. As just seen above, the merger of an adverb and a verb will take place in syntax and not in morphology if the merger involves transparent semantics. This explains the ungrammaticality of *slow(ly)-move* (as opposed to *move slowly*). As for *a slow-moving animal*, following A&N (2004), there is no competition in the non-head position of a compound, and hence the compound is predicted to exist. If one wants to contrast the non-head, i.e. *slow-moving*, with its potential syntactic competitor, i.e. *moving slowly*, the different merger of categories in both cases will suspend competition and the two structures are again predicted to exist. The same explanation can be applied to *fastdrive* and *a fast-driving chauffeur*.

[PV]ᵥ formations[148]

Some authors have included under the class of [PV]ᵥ compounds complex forms which are not the result of compounding but prefixation (Bauer 1983, Carstairs-McCarthy 2002, Katamba & Stonham 2006, Scalise 1984, Selkirk 1982). The P has developed some specific semantics which is used for the creation of complex forms with such meaning and which is different from the semantics of P when found in isolation. This is the case of forms with *out-* like *outachieve*, *outdo*, *outrun*, *outsail*, *outsing*, and

---

[148] Phrasal verbs, i.e. the sum of [V+P]ᵥ, are syntactic objects and hence their omission, but see Katamba & Stonham (2006) for the view according to which phrasal verbs are considered compounds.

outswim, which denote a (potentially) competitive activity and that when attached to intransitive (i.e. unergative) verbs, *out-* transitivizes them by adding an unselected theme argument to the argument structure of the simple verb.[149] Other cases of prefixation involve *over-* with the meaning 'do to an excessive degree' and *under-* with the meaning 'do insufficiently', as in *overcook, overcriticize, overdo, overpolish* and *underachieve, under-dress*, and *underfeed*. If the examples just given were instances of compounding, we would expect the constituents of compounds to have the same meaning as when they are independent words, but this does not seem to be the case. For example, the meaning of *over* as an independent word is not present in verbs like *oversleep, overdo* and *overeat*. These complex forms are then prefixed complex words (see Marchand 1969 and Padrosa-Trias & Markova 2009 for further discussion). (Recall that the same scenario was found for some [PN]$_N$ compounds).

Concerning other complex forms of the type [PV]$_V$, some can be related to synonymous phrasal verbs, V plus P constructions, like *downplay ~ play down*, and *downgrade ~ grade down* (Adams 2001). Others do not have a phrasal counterpart with the same meaning nowadays (e.g. *backslide, backtrack, download, uprate, upstage*), but Berg (1998) convincingly argues that they are the result of inversion from a syntactic phrase originally (e.g. *load down > download$_{N/V}$*). With time the converted form undergoes semantic drift. The upshot is that there seems to be no process of [PV]$_V$ compounding. Put differently, the apparent cases of [PV]$_V$ compounds are either cases of prefixation or of a converted syntactic structure (often associated with semantic drift). That there are no [PV]$_V$ compounds in the language is in agreement with the findings in Booij (2005) and Plag (2003).

[PN]$_V$ formations

Bauer & Renouf (2001) consider cases like (52) compounds, and argue that this type of compound is more productive when the noun is a proper name.

(52) out-Herod, and out-soap.

---

[149] Di Sciullo & Williams (1987) distinguish affixation from compounding by the way the head is related to the non-head. In compounding they are related by theta-role assignment as in *pasta-eater* (although it is not a requirement that needs to be satisfied in every compound), and in affixation they are related by function composition, which means that the head and the nonhead together determine the final argument structure of the complex word. This supports the view according to which *out-* is a prefix, since it visibly contributes to the argument structure of the resulting word. The behaviour of *out-* is exceptional among English prefixes, though, since most of them do not cause any change in the argument structure of the base.

The P *out* in *out-Herod* and *out-soap* has the same meaning as the prefix *out-* in the construction *out*+verb (e.g. *outrun, outsail*) where the prefix denotes "surpass or get the better of (someone or something) in an activity indicated by the base" (Adams 2001: 74; see the [PV]$_V$ formations just discussed above). We have to conclude then that [PN]$_V$ forms are not compounds, but instances of prefixation.

The following table summarizes the results of this subsection. Note that it only includes the verbal compounds available in English (NV, VV, AV) and not the constructions that resemble compounds but are better analysed as a different construction ([PV]$_V$, [PN]$_V$). The compound types which have restricted productivity are included within parentheses, like in Table 2.1.

Table 2.2: Verbal Compounds in English

	SUBORDINATE		ATTRIBUTIVE		COORDINATE	
	VERBAL COMPOUNDS					
	endocentric	exocentric	endocentric	exocentric	endocentric	exocentric
[NV]$_V$	computer-generate (46a)		(chain-smoke) (46b)			
[VV]$_V$	(freeze-dry) (49a)					
[AV]$_V$	(deep-fry) (51)		(soft-land) (51)			

2.3.1.3 Adjectival compounds

We will see below that there are only two types of adjectival compounds: AA and NA compounds. Spencer (2003b: 330) observes that they are "sporadic" and "semantically restricted" and comments that, regarding AA compounds, adjectives cannot usually modify other adjectives (see also Huddleston & Pullum 2002) and concerning NA compounds, they are mostly non-compositional phrases.

First, the right-headed complex forms with an adjective as the second constituent are presented: NA, VA, and AA, of which only the NA and AA forms have been argued to be compounds. Then, it is argued that the PA/AP complex forms are not compounds, nor are the [PN]$_N$ and [VN]$_V$ compounds being used adjectivally. Finally, the VV sequence is discussed and we conclude that it cannot be an instance of adjectival compounds. After surveying the different possible adjectival compounds available in English, we conclude this subsection with a table summarizing the results.

[NA]$_A$ compounds

[NA]$_A$ compounds are all endocentric and can be divided into ATR and SUB compounds, as shown in (53a) and (53b, c, d) respectively. The endocentricity of the compounds comes from the fact that the compound is an adjective, like the constituent in second position, which acts as the formal and semantic head.

(53) a. **ash-blond**, baby-smooth, continent-wide, country fresh, diamond-cut, fire-hot, gift-wrapped (parcels), ice-cold, knee-deep, paper-thin, razor-sharp, shoulder-high, skin-tight, sky-high, star-bright, stone-cold, stone-deaf, and world-wide.

b. **almond-eyed**, bull-necked, chicken-hearted, eagle-eyed, lantern-jawed, pigeon-toed, and wasp-waisted.

c. **sugar-coated**, and chocolate-flavoured.

d. affiliates-led, alcohol-related (incidents), an architect-designed (house), bloodthirsty, a car-dependent (culture), car-dominated (streets), a chauffeur-driven (car), citizens-sponsored, class-conscious, colour-fast, commission-

168

hungry (advisers), computer-matched, a coral-encrusted (wreck), disease inhibitory, drought-ravaged (areas), drug-induced (wisdom), energy-efficient (buildings), flavour-sealed (coffee), foil-packed (coffee), foolproof, germ-resistant, girl-crazy, a glass-fronted (building), guild-laden, guilt-ridden, hand-sewn, hand-written, heat-sensitive, host-specific, issues-oriented, his jail-painted (portraits), leadfree, a London-based (company), a media-shy (financer), moss-covered (rocks), **oil-rich**, paper-bound, poverty-stricken, profits-based, rat-infested, seasick, security-coded (doors), space-born, structure-dependent, sugar-free, sun-baked, tailor-made, time-poor, time-worn, top-heavy, tortoiseshell-framed (spectacles), university-controlled, user-friendly, and water-repellent.

The noun in the compounds in (53a) acts as a modifier of the head adjective, hence the label of ATR compounds. A very common relationship between the constituents of the compound is that of comparison, as in *ash-blond*, which means 'blond like ash'.

It is not clear to us whether colour names like *blood-red* and *bottle-green*[150] should receive the same treatment - i.e. NA compounds with a relation of comparison between the two constituents - or rather should be considered NN compounds similar to the compounds listed in (35b), e.g. *bell jar*. Although our inclination is to treat them as NN compounds, we will leave the question open here. Notice that the two patterns to which these compounds can be assimilated exist independently, so the grouping of colour names like *blood-red* to either pattern will have no effect on the overall compounding patterns of the language (see our treatment of the Catalan counterpart (64a) as an NN compound which might also be applied to English, and see also how compounds like *dark-blue* are treated as an AN compound (cf. the discussion around 55c)).

An attributive relation is also found between the two internal constituents of the compounds in (53b), where the relation of comparison is also prominent: *almond-eyed* means that one's eyes are like almonds in shape. Although superficially the compounds are a sequence of a noun and an adjective, the first noun gives an attribute to the underlying noun on which the adjective is based. The two nouns (e.g. *almond* and *eye*) are, in turn, subordinated to the suffix *–ed*, the head of the compound, which provides

---

[150] Similar colour names are *brick red*, *coal-black*, *grass-green, nut brown*, *onyx-black*, *peacock-blue*, *sea-green*, *sky-blue*, and *snow-white*.

the compound with the adjectival category and the meaning 'having X, provided with X'. In terms of semantics, the suffix attaches to the complex nominal base, as shown by [[almond eye]ed], but phonologically it is added to *eye* (a bracketing paradox). The relation between –*ed* and *almond eye* is one of subordination, hence endocentric SUB compounds. (Recall that according to B&S, they would be exocentric ATR compounds). A requirement for the compounds under analysis appears to be that the second constituent is a body part.

The compounds in (53c) are also endocentric SUB compounds, but they are a bit different. They do not involve a body part as a second constituent and, according to Adams (2001), the relation between the two parts of the compound is ambiguous. They can be understood as 'having a sugar coat' (exocentric SUB following B&S) but also as 'coated with sugar' (endocentric SUB following B&S). The two readings would correspond to two different morphological structures, but they do not have any repercussions on the type of compound. The possibility of having two alternative structures corresponding to different semantics is excluded in the compounds in (53b) and (53d). However, once native speakers have been consulted, it is not clear that there is an extensional difference between the paraphrases 'having a sugar coat' and 'coated with sugar'.

Concerning the compounds in (53d), the noun is an argument or is subcategorized by the adjectival head, as in *structure-dependent*. The noun can also be an adjunct indicating, for example, location as in *his jail-painted (portraits)*. In other words, they are endocentric SUB compounds with the head on the right. In some cases the adjective is underived and then the internal structure is the same as that of the compounds in (53a), but in other cases the adjective is derived and then they have the same internal structure as the compounds in (53b). (Note that in B&S's view, the compounds can be endocentric SUB (e.g. *oil-rich*) and exocentric SUB (e.g. *jail-painted*), the latter due to the mixing of their two levels).

[VA]_A formations

Bauer (1983) observes that adjectival VA compounds exist but that they are rare, and Carstairs-McCarthy (2002: 61) mentions that this type hardly exists, which he attributes to the fact that verbs do not easily combine in compounds, but that it is not difficult to

create new ones, like *sing-happy* 'happy enough to sing' and *float-light* 'light enough to float'. If that were the case, [VA]$_A$ compounds would be endocentric SUB compounds.

(54) diehard, and fail-safe.

However, the reality seems to point in the opposite direction. That is, *fail-safe* is the only oft-cited example which represents the alleged adjectival VA compound, although the result can also be a noun. Regarding *diehard*, Lieber (1983: 255) mentions that the result can either be a noun or an adjective. In short, both *fail-safe* and *diehard* are not prototypical examples of the so-called adjectival VA compound type, which seems to be nonexistent due to the impossibility of creating new forms based on this pattern. We conclude then that there is no adjectival VA compound in English (cf. Booij 2005, Plag 2003, Selkirk 1982), which may be due to the restriction that verbs do not combine with adjectival heads.

[AA]$_A$ compounds
At first sight adjectival compounds of the AA type can be found in all three macro-types of B&S's (2005) classification: CRD (55a, b), ATR (55c) and SUB (55d, 56a, b) compounds.

(55) a. bitter-sweet, blue-green, deaf-mute, devilish-holy, foolish-witty, fortunate-unhappy, harsh-rude, shabby-genteel, phonetic-semantic, sober-sad, social-political, stubborn-hard, and sweet-sour.

   b. cruel-compassionate expression, **French-German cooperation**, the high-low alternation, (one's) humble-surly way, a public-private partnership, and a French-English dictionary.

   c. freezing-cold, **icy-cold**, silky-soft, white-hot, and wide-awake.

   d. barefooted, **blue-eyed**, clear-headed, clear-sighted, flat-chested, good-natured, hard-hearted, long-tailed, long-winded, red-bearded, red-roofed, right-fisted, rosy-fingered, sharp-eared, short-lived, short-sighted, straight-backed, straight-faced, sure-footed, thick-headed, and three-legged.

According to B&S, the compounds in (55a) and (55b) are all CRD compounds, the former being endocentric and the latter exocentric. Concerning the semantics of the former, Lieber (2008) distinguishes 'simultaneous' from 'mixture' endocentric CRD compounds. A compound with a simultaneous relationship between the two constituents is *deaf-mute* and a compound with a mixture relationship between the two constituents is *blue-green*. Given B&S's definition of endocentricity, these compounds are double-headed (although formal marking does not help here because there is none). According to Adams (1973), this type of compound is more common in the literature than in the general vocabulary and, according to Lieber (1983: 255, 260), these compounds can be "readily coined".

Regarding the exocentric CRD compounds in (55b), recall that Lieber (2008) identifies three different relationships, namely relationship, collective and disjunctive, of which we eliminated the last one. The first relation can be exemplified by *French-German cooperation*; the second one by *a cruel-compassionate expression*. Recall also that Bauer (2008) also distinguished two subtypes of compounds: translative (*a French-English dictionary*) and co-participant (*French-German cooperation*).

Recall that our understanding of compounding leads us to deny the existence of CRD compounds (see also Adams 2001: 97). For us, the forms in (55a) are just phrases with asyndetic coordination. While the CRD relation may not be visible in forms like *bitter-sweet* and *blue-green*, which may have come to denote for some speakers a kind of flavour and a new colour respectively (as a result of mixing the two coordinated elements), in other cases the CRD relation is clearly visible: e.g. *shabby-genteel* is understood as genteel but shabby, and *social-political* as involving both social and political aspects. Some forms are treated by speakers as not making much sense: e.g. *devilish-holy*, *foolish-witty*, *fortunate-unhappy* and *harsh-rude*. By their very nature, adjectives need to be attributed to some entity, which explains why the nonsensical coordinated adjectives are rescued by placing a noun outside them: e.g. *devilish-holy fight/fray, foolish-witty love, fortunate-unhappy thing, harsh-rude tongue*. Note that it is also due to their (contradictory) semantics that the two adjectives cannot easily enter into a CRD relation; the insertion of a noun makes such a relation possible.

Similarly, the coordinated forms in (55b) do not make much sense unless they are predicated of a noun outside the coordinate structure: *??public-private* vs. *a public-private partnership*. In short, it seems that an AA sequence can easily be interpreted as a

coordination when it is inserted in the nonhead position of a compound which has a head noun whose inherent semantics licenses a coordinate relation. For example, *partnership* semantically involves two or more people, a requirement which is satisfied in *a public-private partnership*. In other words, the forms in (55b) are compounds by virtue of having a noun outside the coordinate structure. They are endocentric compounds with the noun being the formal and semantic head. What is not so clear is whether they are SUB or ATR compounds: *a public-private partnership* is a partnership *between* public and private organisations (SUB) but *a cruel-compassionate expression* is an expression which *is* cruel but compassionate (ATR). The issue of the SUB/ATR division will be taken up in the discussion section. They will provisionally be placed under SUB compounds in Table 2.1 (and not in Table 2.3, which contains adjectival compounds) and will be treated as nominal AN compounds with a coordinate adjectival phrase in the non-head position.

The compounds in (55c) are endocentric ATR. They consist of a sequence of two adjectives: the first one modifies the second one, which is the head. For example, *icy-cold* indicates that it is as cold as ice. They exemplify the same pattern we found for adjectival NA compounds like *ash-blond* (53a). Like the compounds in (55a), Lieber (1983: 255, 260) also considers that AA compounds like *icy-cold* and *wide-awake* are "readily coined". By contrast, according to Huddleston & Pullum (2002: 528), adjectives cannot be modified by other adjectives, except for a few cases: *icy cold*, *freezing cold*, *red-hot*, *boiling hot*, *scalding hot*, of which they note that "These expressions have the character of fixed expressions (…) this is not a productive construction" (p. 550, fn. 7). Notice that some modifying adjectives are based on a noun: *icy<ice*, *silky<silk* while others are based on a verb: *freezing<freeze*. If Huddleston & Pullum's view is correct, then we have reason to believe that forms like *bluish-green*, *dark-blue*, *light-green*, *pearly-grey* and *snowy-white*, which could initially be taken as AA compounds, should be considered AN compounds.

The compounds in (55d) are endocentric SUB, with the same structure as the adjectival NA compounds which we have seen above (cf. *almond-eyed* (53b)), but with the compounds at hand there is no relation of comparison between the two internal constituents but a copulative relation. For instance, in *blue-eyed*, the eyes are blue, which means that *blue* modifies the noun *eye* underlying the second adjective. The head

of the compound, the ornative derivational suffix *–ed*, attaches to *eye* formally, resulting in *eyed*, but semantically attaches to [blue eye] giving it the meaning of 'having X (=blue eyes)' and also providing it with the adjectival category (the result being a bracketing paradox, cf. Spencer 1991). All the compounds involve a body part (or something related to the human body like *wind*, *nature*, *life*) as the second internal constituent, which is inalienably possessed by the noun the compound as a whole modifies: *a [[three$_A$ leg(g)$_N$]ed]$_A$ table* (the table necessarily has legs or something similar to legs) vs. **a [[two$_A$ skirt$_N$]ed]$_A$ woman* (the woman does not necessarily need to have skirts). In short, in addition to the ATR relation between *blue* and *eyes*, there is a SUB relation between the suffix *–ed* and *blue eye*. While B&S (2005) take the innermost relationship to classify the compounds as ATR, we think that the ATR relation is subjected to the outermost SUB relation, and hence we consider that the compounds are SUB with an ATR relation inside. Compounds very similar to the ones in (55d) are given in (56):

(56)   a. **free-spirited** (culture), and a low-powered (airgun).

   b. two great-coated (figures), our **light-industrial** (heritage), and his many-pocketed (fishing vest).

A difference between the compounds in (55d) and those in (56) is that the second constituent of these compounds is not a body part. As for the compounds listed in (56b), the noun outside the compound does not need to inalienably possess the entity which is suffixed with *–ed*: in *two great-coated (figures)*, figures do not necessarily have coats.[151]

[PA]$_A$ formations

Under the heading [PA]$_A$ compounds, some authors like Carstairs-McCarthy (2002) have included complex forms which are the result of prefixation rather than compounding. As discussed earlier (e.g. in the case of some nominal [PN]$_N$ compounds), the Ps *over* and *under* have developed a specific meaning, which is

---

[151] Some of these examples seem to fit in with Spencer's (1991: 417) generalization that the second constituent typically is an external and internal body part or a clothing part. This still does not cover all cases (e.g. *light-industrial (heritage), five-pointed*).

different from the meaning of the P when it is found in isolation, suggesting that there is a bound prefix form besides a free form. For example, *over-* intensifies the property denoted by the adjectival base, as in *over-aggressive*, *over-confident*, *over-concise*, and *overripe*, a meaning not found when *over* functions as a free form (e.g. *The plane flew over the city*) (see Lieber 1983: 261, fn. 16, for a similar view). As for the rest of the apparent adjectival PA compounds, Berg (1998) shows that they are the result of inversion from a syntactic combination (e.g. *built in > inbuilt$_A$*). While some PA formations have developed a specialised meaning different from their syntactic source (*come up > upcoming*) due to the passage of time, other complex PA words can still be semantically associated with the syntactic phrase from which they are derived (e.g. *speak out > outspoken*). The fact that one cannot put any P and A together to form a compound supports the hypothesis that there are no adjectival PA compounds in English. In short, alleged cases of [PA]$_A$ compounds are either the result of prefixation or the result of a verbal phrase undergoing conversion into an adjective.

[AP]$_A$ formations

Forms like *strung out* and *worn out* also develop out of phrasal verbs. The semantic link between some adjectival forms and their syntactic source is still visible in some cases (e.g. *wear out* and *worn out* are both related to tiredness and exhaustedness) but opaque in other cases (*string out* means 'to spread out', but *strung out* means 'to be addicted to a drug'). In other words, the source for the alleged compound under consideration is a phrasal verb, which is converted into an adjective later on (cf. Berg 1998). Like in the previous case (i.e. [PA]$_A$ formations), there is no [AP]$_A$ compounding process.

[PN]$_A$ formations

As we saw in the subsection on nominal compounds, P+N is a possible combination for a compound noun (e.g. *outpost*, *underarm*, *out-tray* (cf. 39)). The compounds listed in (57) seem to have the same internal structure as the compounds in (39b), an example of which is *underarm*, and as the forms presented as potential prepositional compounds in subsection 2.3.1, an example of which is *offstage*. In all cases, the internal argument of the P is satisfied by the noun. We understand that the result of merging a preposition with a noun can be adjectival when it has undergone semantic drift (to some degree) and the meaning is not totally transparent. Some evidence for them acting as adjectives

comes from their placement in comparative contexts: *They live in a very downmarket neighbourhood* (Carstairs-McCarthy 2002: 65; cf. Wasow 1977).

(57)   before-tax (profits), downmarket, in-house, off-shore, upmarket, upscale, and with-profits.

It seems that once P+N words have been incorporated into the language, they are subject to historical accidents: some words may have remained faithful to their category (e.g. *offstage$_P$)*, some others may have developed a nominal use (e.g. *afterbirth$_N$* in (39b)), others may have developed adjectival and nominal uses (e.g. *underarm$_{N/A}$, underground$_{N/A}$*) and still others may have ended up being used as adjectives only. Note that not all the forms in (57) exemplify the latter development: e.g. **a very off-shore account* (contra Carstairs-McCarthy 2002). *Off-shore* seems to have the same role as the PP *off the shore*, with the reduced form probably being a preposition. Since [PN]$_A$ forms cannot be created spontaneously, the forms in (57) cannot be taken as compounds. Notice that the existence of such words depends on a conversion process, which is unpredictable.

[VN]$_A$ formations

The words in (58) seem to have their source in nominal [VN]$_N$ compounds (cf. 37b: *cutthroat, telltale*). The noun is the internal argument of the verb, and a zero-affix satisfies its external argument and gives the nominal category to the compound. As in the previous case, we believe that once such forms have been in the language for some time, they can deviate from their original meaning and can subsequently develop other usages, like an adjectival one. Consider the following cases: *at breakneck speed, a cutthroat razor*, and *kick-arse attitude*. These examples show that *breakneck, cutthroat* and *kick-arse* cannot be interpreted literally. For example, *breakneck* does not literally mean 'to break somebody's neck' but 'very fast'. Similarly, *cutthroat* is interpreted as 'cruel' and *kick-arse/kick-ass* as 'very tough, aggressive'. Since the existence of [VN]$_A$ forms seems to depend on the existence of [VN]$_N$ compounds and the categorial change (N→A) is not systematic, the conclusion is that [VN]$_A$ words do not constitute a compound type. Semantic drift seems to be a requirement for VN forms to be able to be used adjectivally.

(58)   breakneck, catchpenny, cutthroat, kick-arse, lacklustre, and telltale.

Note that nowadays most of these forms are used only adjectivally (e.g. *lacklustre*) and in fixed expressions (e.g. *at breakneck speed*). This fact may be attributed to the fact that the entities that these forms may have denoted originally are no longer relevant in our world (e.g. *cutthroat* designating a criminal). One might propose another analysis, namely that the forms under study are converted VPs, which would explain why such forms seem to lack a nominal counterpart (at least nowadays). However, we believe such a proposal cannot explain why native speakers cannot come up with new forms since the pattern which would underlie the process is an active one in the language (a verb plus its internal argument).

[VV]_A formations

We have found only the two examples in (59) that conform to the [VV]_A pattern, which may be indicative of the low presence of this compound in the language if it exists at all.

(59)   make-believe, and wash-wear.

As already noted in the subsection on verbal compounds of the VV type, the source of *make-believe* seems to be a sequence of two verbs in syntax, with a relation of subordination between them (*make* being the head). Losing its transparency, this verbal sequence has acquired an adjectival function and can now be used in sentences like: *She has a make-believe friend* and *His story is all make-believe*. Regarding *wash-wear*, it looks like a lexicalized phrase of *wash-and-wear*, a syntactic phrase as well. This leads us to conclude that there are no real cases of adjectival VV compounds. Examples in which the forms in (59) are used in context are given below:

(60)   a. "Discuss ways to tell the difference between books that tell make-believe
       stories and books that tell real facts."
       b. "Tom Sawyer outfits were not only stylish, but were washwear meaning they
       could be easily laundered."[152]

---

[152] The examples in (60) are the result of a Google search.

This concludes our survey of adjectival compounds and the results from this subsection are given in Table 2.3. Like Tables 2.1 and 2.2, the present one only includes the complex words that have been argued to be compounds (NA, AA) and the low productivity of some compound types is indicated by means of parentheses.

Table 2.3: Adjectival Compounds in English

	ADJECTIVAL COMPOUNDS					
	SUBORDINATE		ATTRIBUTIVE		COORDINATE	
	endocentric	exocentric	endocentric	exocentric	endocentric	exocentric
[NA]_A	almond-eyed (53b) sugar-coated (53c) oil-rich (53d)		(ash-blond) (53a)			
[AA]_A	blue-eyed (55d) free-spirited (56a) light-industrial (56b)		(icy-cold) (55c)			

2.3.2 Catalan

After presenting some sources on which the Catalan survey is based, this subsection reveals our position with respect to several phenomena which will be relevant in the survey of Catalan compounding: the nature of prefixes and adverbs ending in –*ment* (derivation vs. compounding), the role of stress in compounding and the constituting categories of compounds.

The survey of Catalan compounding includes the following sources: Adelman (2002), Badia (1962), Brunelli (2003), Cabré & Rigau (1986), Duarte & Alsina (1986), Fabra (1956), Ferrater (1981), Gavarró (1990b), Gràcia (2002), Gràcia & Fullana (1999, 2000), Grossmann (1986), Klingebiel (1988), Mascaró (1986) and Moll (1952, 1975). The smaller number of studies devoted to compounding in Catalan than those dedicated to English compounding has been reinforced by looking at other Romance languages like Spanish, Italian, Portuguese and French (e.g. Benveniste 1974, Bisetto & Melloni 2008, Bisetto & Scalise 1999, Bok-Bennema & Kampers-Manhe 2006, Buenafuentes 2001-2002, 2007, Clements 1992, Contreras 1985, Fábregas & Scalise 2008, García Lozano 1978, Gil Laforga 2003, 2006, Kornfeld 2003, Lang 1992, Oniga 1992, Piera & Varela 1999, Rainer & Varela 1992, Sánchez López 2003, Scalise 1992, Val Álvaro 1999, Varela 1989, 1990, 1999, 2005, Villalva 1992, Zwanenburg 1992b, a.o.).

The status of prefixes as part of derivation or compounding has been much debated (see Cabré 1988 for a review of the early literature on the topic). Some authors consider them to be part of compounding without making any distinction between subtypes of prefixes (e.g. Fabra 1956, Moll 1975); others treat prefixes as part of derivation (e.g. Badia 1962, Buenafuentes 2007, Cabré 1994, Cabré & Rigau 1986, Gràcia et al. 2000); and still other authors like Kornfeld and Saab (2003), on the one hand, and Cabré (2002) and Gràcia (2002), on the other, make a distinction among prefixes.

Understanding morphology as word syntax à la Lieber (1992) and argument structure in the sense of H&K (1998), Kornfeld and Saab (2003: 239) argue for the following view:

(62)   "the traditional notions of derivation and compounding are not primitive but derived concepts, inasmuch as what is relevant is the operation involved in generating the lexical item (conflation or merge), which depends in its turn on the component of the grammar at work (word syntax or proper syntax)".

Kornfeld and Saab provide three different processes to account for complex words formed by a prepositional prefix followed by a N, an A and a V in Spanish. P+N formations like *sin+vergüenza* (without+shame) 'rascal' would be the result of merging two heads in proper syntax with the consequence of creating a complex head, a process which they call syntactic compounding. With time P+N forms can become less transparent and need to be listed in the lexicon. Concerning P+A formations, they focus on the Greek and Latin prefixes *pro-*, *inter-* and *anti-* and conclude that when they attach to relational adjectives (e.g. *anti+gubernamental* (anti+governmental)), they are cases of conflating affixes in word syntax, the result being bracketing paradoxes: semantically the prefix modifies the underlying noun but formally it attaches to the derived adjective. Finally, P+V constructions like *sobrepasar* (over+pass) 'exceed' are the result of conflating words in the lexicon (i.e. word syntax), which they call morphological compounding. To our understanding, if P+N forms were the result of syntax with subsequent listing, they cannot be treated as compounds nor can P+A forms, which following Kornfeld and Saab's argumentation should be treated as cases of derivation. P+V forms would be the only ones which can be taken to be compounds: they are the result of combining two free forms, which in our view would take place in the morphological component. In search of a uniform treatment for all P+X forms, Kornfeld and Saab's view cannot be adopted.

Cabré (2002: 739) and Gràcia (2002: 782-786) make a distinction between stressed prefixes and non-stressed prefixes. The former are considered as forming part of compounding and include those prefixes that have an independent counterpart in the language (i.e. preposition) as well as those prefixes that come from Greek and Latin and used to be prepositions or adverbs in these languages. The latter (i.e. non-stressed prefixes) are seen as derivational affixes and include those prefixes that have no independent counterpart in syntax. In this thesis we will adopt Cabré and Gràcia's view of prefixes, with some revisions. That is, we will consider compounds those forms whose first element is a prefix which has a counterpart in syntax with the same meaning. Notice that this greatly reduces the number of potential compounds, given that all Greek- and Latin-based prefixes will be excluded: they do not have an independent counterpart. In addition, if the prefix in the would-be compound has developed specific semantics different from the semantics associated with its independent counterpart, then it will not be considered a case of compounding, but rather of derivation, as we have

already seen for English (for example, recall that the meaning of –*over* in *overcook* is no longer locative, as in the preposition *over*, but it indicates degree).[153] As Varela & Martín (1999) have concluded for Spanish, there is a general pattern for prefixes with a locative meaning to adopt and specialize in other semantic interpretations.

Similarly, the question of whether adverbs ending in –*ment* '-ly' belong to compounding (e.g. Moll 1952, Mascaró 1986, Wheeler 1977) or derivation (e.g. Cabré & Rigau 1986, Cabré 2002) is still not settled in the literature.[154] The complex forms have an adjective in the feminine form as the first element and –*ment* as the second element. On the one hand, if –*ment* were a derivational suffix, it would follow an inflectional suffix (that of the feminine marker), which would go against the usual order of derivation coming first and inflection coming later. Note also that the complex word has two stresses, a phenomenon which is traditionally associated with composition although should no longer be, as will be seen below. That is, the presence of two stresses should not be taken as an indication of identifying a compound. On the other hand, if *ment*-adverbs were a case of compounding, it would constitute the only type whose resulting category is adverbial, and it would be a counterexample to Cabré & Rigau's (1986) generalization that the resulting categories of compounds in Catalan are nouns, adjectives and verbs. More importantly, the semantics of *ment* in the alleged compound cannot be associated with the independent word *ment* 'mind' in Catalan (which is in turn derived from the Latin noun *mens*, *mentis*).[155] Furthermore, the category of the compound is an adverb, which is an unexpected result if we consider that the first element is an adjective and the second one is supposedly a noun (if one wants to link it to the free element *ment* 'mind'). That complex forms ending in –*ment* are not compounds is further corroborated by the fact that the non-head in such complex forms can have its argument satisfied in external syntax, a fact unknown to compounding. Torner (2005: 123) provides the following examples for Spanish, which are also applicable to Catalan:

(63)   a. simultáneo a esa acción

---

[153] This also applies to some Greek- and Latin-based combining forms (cf. subsection 2.2.1) which behave as prefixes: most of these prefixes do not have the same semantics as the corresponding combining form (e.g. *super-*, *hiper-*) and are thus excluded from our study of compounds. (See e.g. Buenafuentes 2007: 369-372, who treats such forms as the result of grammaticalization).

[154] To our knowledge, the English counterpart does not present the same problem as the Catalan (or Romance) –*ment*: the ending –*ly* has not controversially been taken as a compounding element.

[155] Buenafuentes (2007: 20) observes that already in late Latin the semantics of –*ment* had changed from the meaning of 'mind' to 'in such a manner'. Despite this fact, Baker (2003: 234f) treats –*ment* as a noun.

'simultaneous to this action'

b. simultáneamente a esa acción

'simultaneously to this action'

This concludes our short discussion about the status of adverbs ending in –*ment*: they are not compounds and thus will not be considered in what follows. Torner argues for Spanish that –*mente* is a phrasal affix that attaches to APs and so does Gavarró (1990b) for Catalan –*ment*, a proposal which we view as plausible.

Counting the number of stresses of a word was traditionally used as a test to distinguish derived words from compounds: derived words typically had one stress while compounds had two. However, authors like Mascaró (2002: 116-117) and Prieto (2003) have recently shown that the first element of compounds is in fact produced and perceived as if it had no stress. For example, Prieto's (2003) production and perception experiments show that there are no robust acoustic correlates for the existence of a secondary stress in compounds. Padrosa-Trias (in press, b) concludes that "(…) what seems to distinguish derivation from compounding is not the number of the stresses the word bears but the presence or absence of vowel reduction.[156] Accordingly, derivation is typically associated with vowel reduction, while compounding is not". Vowel reduction, nonetheless, will not be taken into account because it is subject to variation depending on the dialect. Rather, the semantic specialization of a form will help us determine whether it is part of compounding or derivation, as was the case of the prefix *over-* in English, for example.

As in the survey of English compounding, there are no prepositional compounds of the P+P type in Catalan. The only cases which could be treated as such are lexicalizations of two prepositions occurring together: *des de* (from of) 'from', *envers* (in wards) 'towards', *fins a* (until in) 'until', *per a* (for in) 'for', and *per contra* (for contra) 'on the contrary'. Like in English, there are forms made up of a preposition and a noun which could be taken as prepositional compounds, but the result also seems frozen and no new creations are possible: e.g. *darrere* is analysed as [de$_P$+rere$_{NP}$] (of+back) 'behind' in Bartra & Suñer (1992: 51), a work which also provides similar

---

[156] Vowel reduction can be illustrated with the contrast between *poma* vs. *pometa* (apple vs. apple-DIM) which are pronounced as /'p□m↔/ vs. /pʏ'mEt↔/: the first vowel of *poma* undergoes vowel reduction in its diminutive form (derivation). By contrast, there is no vowel reduction in the complex forms that we have defined as compounds: e.g. *espanta+sogres* (scare+mothers.in.law) 'party blower' is pronounced as /↔s'pɑnt↔'s□gr↔s/: the second vowel of *espanta* does not undergo vowel reduction in the compound.

forms from other Romance languages (e.g. Spanish, Galician, Italian). The frozen status of the existing forms and the impossibility of creating new forms leads us to conclude that there are no prepositional compounds of the P+N type in Catalan (and probably in Romance in general). As will be seen in each of the three subsections that follow, the resulting category of a compound cannot be prepositional, but prepositions can form part of the constitutive elements of compounds (although this view is not accepted across the board: Cabré & Rigau 1986). Finally, recall that although we think that adverbs are a derived category (the base form being the adjective; cf. Emonds 1976, Gavarró 1990b: 138f, a.o.), we keep using the labels 'adjective' and 'adverb' for convenience (and also in line with traditional studies of morphology). As will be seen shortly, the categories participating in compounding are N, V, A, and P, with the category A standing for both adjectives and adverbs.

The next three subsections include our survey of compounding in Catalan: subsection 2.3.2.1 contains the nominal compounds, which is followed by the verbal compounds (subsection 2.3.2.2) and the adjectival compounds (subsection 2.3.2.3). Recall that Catalan compounds can be spelt as one word, as two words or hyphenated and that the two elements of the compound are joined with a '+' sign when they are spelt as one word (cf. see footnote 4 in chapter 1 for the format followed in the gloss).

2.3.2.1 Nominal compounds

There is quite a wide range of nominal compounds, although most of them cannot be regarded as productive nowadays. Three types of nominal compounds can be distinguished, all of them having a noun as their second member: NN, VN and PN. NN compounds predominate over the others: apparently they can instantiate all three possible grammatical relations between the two constituents (SUB, ATR and CRD), although in practice not all relations are productive and CRD compounds cannot be considered compounds in the sense we understand them (i.e. produced by the morphological component). The only possible relation in VN compounds is the SUB one, a compound which is regarded by some authors (e.g. Mascaró 1986) as the most productive one. There are also a few PN compounds, although not everybody agrees that they should be included in a classification of Catalan compounds (e.g. Cabré 1994). In addition to NN, VN and PN compounds, there are four other nominal types in which the second element is not a noun: VV, NA, NV and PA. None of them will be argued to be compounds. This subsection of nominal compounds concludes with some discussion

of certain expressions (i.e. AN, QuantN, NA, N prep N) whose status has proved to be controversial: while some authors regard them as (synaptic/syntagmatic) compounds, others have not included them in their classifications of compounds. We will adopt the latter position: we will consider such forms lexicalized phrases, and hence not part of Catalan compounding. The results of this subsection are gathered in Table 2.4.

[NN]$_N$ compounds

Nominal compounds of the NN type all seem to follow Meyer's (1993) requirement that the non-head make the meaning of the head more specific: the non-head should not denote a superset of the head noun or be a necessary part of the head. As will be seen below, all NN compounds satisfy such a requirement: e.g. in *faldilla pantaló* (skirt trousers) 'skort' (64a), the non-head noun *pantaló* in second position specifies a type of the head noun, *faldilla*. In other NN compounds, the compound as a whole refers to the entity denoted by the second noun (the non-head), which is a type of object denoted by the first noun: *porc senglar* (pig+wild boar) is a *senglar* 'wild boar' (a type of pig), *pala+fanga* (spade+pitchfork) is a *fanga* 'pitchfork' (a type of spade), *terr(a)+argila*[157] (earth+clay) denotes *argila* 'clay' (a type of earth). The fact that the non-head denotes an entity which is a sub-type of the entity denoted by the head makes the head redundant and sometimes leads to its deletion, which could be the case of *fanga* and *argila* (see below).

[NN]$_N$ compounds can initially be divided into ATR (64), CRD (65) and SUB (66-67) compounds. ATR compounds are all endocentric and they have been divided into two groups. While the forms in (64a) can be considered clear cases of compounds, some of the forms in (64b) may be treated as simplex forms by some speakers.

(64) a. barca cisterna (boat tanker) 'tanker boat', blau cel (blue sky) 'sky blue', blau turquesa (blue turquoise) 'turquoise blue', cafè teatre (café theatre)[158], cartró pedra (cardboard stone) 'papier-mâché', cine-club (cinema+club), ciutat dormitori (town dormitory) 'dormitory town', cotxe bomba (car bomb) 'car

---

[157] In some compounds, the final vowel of the first element coincides with the first vowel of the second element: both are [ə]. In these cases, the two vowels are fused into one, which we have signalled by including the first coincident vowel within parentheses, as in *terr(a)+argila* (earth+clay) 'clay'. In some other cases, the two vowels are spelt identically but are not the same: the first one is [ə] and the second one is [á], as in *mal(a)+ànima* (bad+soul) 'a cruel, heartless person'. In these cases, the unstressed vowel is elided, which we have also signalled by including it within parentheses (see e.g. Wheeler 1977, Recasens 1993, Prieto 2004 for discussion on this point).

[158] When the word-by-word gloss is transparent, the corresponding translation is not given.

bomb', cotxe escombra (car broom) 'chase/support car', **faldilla pantaló**[159] (skirt trousers) 'skort', figa+flor (fig+flower) 'early fig' and 'weak character', gos llop (dog wolf) 'wolf hound', gris perla (grey pearl) 'pearl grey', groc canari (yellow canary) 'canary-yellow', home anunci (man advert) 'ad man', home aranya (man spider) 'a spiderman', hotel apartament (hotel apartment) 'apartment hotel', lliri+jonc (lily+rush) 'sword lily', malva poma (mauve apple) 'apple mauve', peix martell (fish hammer) 'hammer fish', paper moneda (paper coin) 'paper to be used like coins, banknote', pis patera (flat + little boat)[160] 'a small flat where many immigrants live together', targeta postal (card postal) 'postcard', tren hotel (train hotel) 'hotel train', vagó llit (wagon bed) 'sleeper', vagó restaurant (wagon restaurant) 'dining car', **verd oliva** (green olive) 'olive-green', verd poma (green apple) 'apple green', and verd turquesa (green turquoise) 'turquoise green'.[161]

b. **blau gris** (blue grey) 'greyish blue', blau verd (blue green) 'a greenish blue', porc senglar (pig+wild boar) 'wild boar', pala+fanga (spade+pitchfork) 'pitchfork', terr(a)+argila (earth+clay) 'clay', and verd blau (green blue) 'bluish green'.

The endocentric ATR compounds in (64a) are alive in the language, which explains why we can create novel instances: *un jardí museu* (a garden museum) 'a garden which is also a museum', *una casa alberg* (a house youth.hostel) 'a house which functions like a youth hostel', *una maleta maletí* (a backpack briefcase) 'a backpack which resembles a briefcase', *una piscina aquari* (a swimming.pool aquarium) 'a swimming pool which may have fish like an aquarium', *un boli escopeta* (a pen gun) 'a pen which can function as a gun'. This is why we disagree with Gavarró's (1990b: 186) statement that NN compounds of this type are "rare and unproductive". Of a similar opinion is Martí

---

[159] Authors like Varela (2005) and Buenafuentes (2007) treat compounds like *falda pantalón* (skirt trousers) 'skort', the Spanish counterpart of the Catalan *faldilla pantaló*, as a CRD compound. For us, the first noun is the semantic head: the compound is a type of skirt, which happens to have some properties associated with trousers, namely a pair of legs. That is, we do not understand the compound as an entity which is equally characterized, in terms of semantics, by the two elements of the compound (cf. see the discussion around the examples in (65c)).

[160] *Patera* 'little boat' is wrongly used to refer to any boat used by immigrants who are from South Africa and try to reach Spain illegally. This misconception has given the compound *pis patera*.

[161] Colour names like *blau clar* (blue light) 'light blue', although similar to compounds listed in (64), are not compounds themselves. Note, for example, that *clar* can undergo degree modification: *un blau [molt clar]* (a blue [very light]) 'a very light blue' (see Bosque 1989: 114-118, Varela 2005: 82-83)

(2002: 1319-1320), who observes that such forms have limited productivity. We agree that not every NN compound is possible, but we attribute this fact to the speakers' encyclopaedic knowledge not permitting the felicitous combination of some nouns. We further disagree with Gavarró when she observes that these compounds are lexicalized phrases. If they really were lexicalized phrases, new formations would not be possible and functional material would be expected to intervene between the two nouns, contrary to reality. Precisely the absence of syntactic principles leads us to believe that these formations are regulated by morphological principles, and hence should be considered compounds. (We will not enter into the distinction between appositions and compounds, but see e.g. Rainer & Varela 1992: 118-120, Val Álvaro 1999: 4778-4783).

According to Benveniste (1974: 147), the entity the compounds refer to belongs to two classes of objects, one which the entity belongs to naturally, the second one the entity is identified with only figuratively. To illustrate the point, *un home aranya* (a man spider) 'a spiderman' is a man who has some similarities to spiders. Among the ATR compounds in (64a), some authors identify different types. For example, Val Álvaro (1999: 4784) distinguishes classifying compounds and qualitative compounds. In classifying compounds the second noun classifies the first noun, as in *vagó llit* (wagon bed) 'sleeper', and in qualitative compounds the second noun qualifies the first noun, as in *ciutat dormitori* (town dormitory) 'dormitory town'. This distinction is overlooked here, since it does not add to the purpose of the chapter.

Note that some of the compounds included in (64a) could also be treated as SUB compounds (cf. Buenafuentes 2007). In addition to the ATR reading, according to which the first noun is seen as having a property (or some properties) of the second noun, compounds like *barca cisterna* (boat tanker) 'tanker boat', *vagó llit* (wagon bed) 'sleeper', and *vagó restaurant* (wagon restaurant) 'dining car' can also be interpreted as the first noun containing the second one, thus deriving the SUB reading. We find the first reading more plausible (e.g. in *vagó restaurant* (wagon restaurant) 'dining car', we interpret that the wagon functions like a restaurant), which explains why these compounds are listed under ATR compounds and not as SUB compounds (but see section 2.4 where the SUB-ATR distinction is further discussed and clarified).

From the discussion above, we gather that the compounds are left-headed semantically with the second noun assigning an attribute to the first noun (ATR relation): *peix martell* (fish hammer) 'hammer fish' is a type of fish which resembles the shape of a hammer. We follow Gràcia (2002: 820) and Pérez Saldanya et al. (2004:

255) in that we take the names of colours like *blau turquesa* (blue turquoise) 'turquoise blue' to be nouns, but for a different view, see Badia (1962: 383), Duarte & Alsina (1986: 153) and Mascaró (1986: 72), who regard them as adjectives, or Zwanenburg (1992b: 227), who treats the French counterparts as nominalized adjectives. (For discussion of colour names in Spanish, see e.g. Gallardo 1981, González Calvo 1976, Bosque 1989, among others). The way inflection works for such colour names may be taken as evidence for treating them as nouns and not as adjectives. In general, they are invariable. When the compound has a predicative value, the noun may sometimes agree in number but not in gender: *jerseis blau+(s) turquesa* (jersey-PL blue(-PL) turquoise), **faldilla blav+a cel* (*skirt.FEM blue-FEM sky) (example from Gràcia 2002: 821; but see Picallo 2002: 1649 and Pérez Saldanya et al. 2004: 252 where agreement in number is also disallowed). When the compound has a nominal role, then the head tends to be inflected: *Són tres blau+s turquesa diferents* ((They) are three blue-PL turquoise different-PL) '(They) are three different turquoise blues' (example from Gràcia 2002: 821). If we follow the canonical postnominal position of the adjective in Catalan, which is the position in which the adjective has a restrictive role (cf. Picallo 2002: 1655), we will conclude that the colour names involved in the compound under study must be nouns and not adjectives. If *blau* in *blau turquesa* (blue turquoise) were an adjective, it would be in a prenominal position since *turquesa* can only be a noun, a fact which goes against the canonical placement of the adjective. In addition, they also behave as nouns in the sense that they do not allow degree modification by *very*, as in *Les flors del jardí eren (*molt) blau turquesa* (The flowers in the garden were (*very) blue turquoise), or the superlative, as in *El jersei era blav(*íssim) turquesa* (The sweater was blue(*est) turquoise). Thus, we conclude that forms like *blau turquesa* are nominal NN compounds (and not AN compounds). Other Romance languages show the same results as those illustrated here for Catalan (see e.g. Clements 1992 for Spanish, Villalva 1992 for Portuguese).[162] As for the rest of compounds in (64a), plural marking is placed on the semantic head: *goss+os llop* (dog-PL wolf), *cotxe+s bomba* (car-PL bomb), although some speakers accept plural marking on the two elements for some compounds: *goss+os llop+s* (dog-PL wolf-PL), *cotxe+s bombe+s* (car-PL bomb-PL). Although plural marking can be accepted on both elements of the compound sometimes and the nominal category of the compound could come from either element, note that the diminutive

---

[162] Note that in Italian similar compounds like *rosso fuoco* (red fire) 'fire-red' are treated as [AN]$_A$ compounds by Gaeta (2006) and as [AN]$_N$ compounds by Bisetto (2004).

suffix can only attach to the first noun, i.e. the semantic head*: peix+et espasa* (fish-DIM sword) 'small sword fish'. Similarly, when there is a difference in gender between the two compounding elements, the gender of the word as a whole is determined by that of the first element: *[un [peix]$_{masc}$-[espasa]$_{fem}$]$_{masc}$*. We then conclude that the first noun is the head both formally and semantically, hence the endocentricity of this compound type. When the compound becomes lexicalised like *figa+flor* (fig+flower) 'weak character', the compound is treated as a simplex word and plural marking goes to the end of the word. (The same compounding type is found in other Romance languages, but note that not all apparently identical NN sequences can be treated as compounds. For example, Grandi 2009 discusses Italian NN sequences like *riunione fiume* (meeting river) 'very long meeting' which at first sight look like the forms listed in (64a). They cannot be treated as the compounds under discussion but as syntactic phrases: the second noun behaves like an adjective semantically and formally. That is, it has a modification adjectival function and can itself be modified by degree modifiers like *very* and *more*, and can be used in the superlative, among other properties typical of adjectives. This behaviour is not found among the forms in (64a).)

To this compound type, we could add the compounds discussed in the introduction to NN compounds and listed in (64b), namely *porc senglar* (pig+wild boar) 'wild boar', *pala+fanga* (spade+pitchfork) 'pitchfork', and *terr(a)+argila* (earth+clay) 'clay', although they behave differently in some respects. They are also ATR compounds. The ATR relation between the two nouns makes sense in either direction semantically: e.g. in a *pala+fanga* (spade+pitchfork) 'pitchfork', the spade is a pitchfork and a pitchfork is a spade. However, following Meyer's (1993) semantic restriction (see subsection 2.3.1.1: nominal NN compounds in English), we conclude that the ATR relation is as indicated in the first case (i.e. the second noun is attributed to the first noun), in which one characterizes a general object by specifying a subtype of it. Semantically, it seems that these compounds are double-headed. That is, the compound can be understood as a hyponym of either constituent: a *pala+fanga* (spade+pitchfork) is both a spade and a pitchfork, the latter being more specific. Formally, the compound gives mixed results. In the case of *porc senglar*, the plural is placed on both elements (*porc+s senglar+s*), but the diminutive marking can only be used with the first noun (*porqu+et senglar*), thus indicating that the first noun is the real formal head. *Porc senglar* is then no different from the compounds in (64a). As for the other two

compounds, plural and diminutive marking goes to the end (e.g. *pala+fangu+es*, *pala+fangu+eta*: spade+pitchfork-PL/DIM), which is probably due to the speakers' perception of them as simplex words. If this view is correct, these forms cannot be considered as forming part of the inventory of Catalan compounding, and in fact *terr(a)+argila* is no longer attested in some dictionaries (e.g. GDLC and *Diccionari Institut d'Estudis Catalans*: DIEC), which points to the fact already noted at the beginning of this compound type. That is, NN compounds can be reduced to one noun if the information provided by the other is seen as redundant, which seems to be the case of *terr(a)+argila*, and *pala+fanga* may be on its way (or already it is for some speakers consulted, who only talk of *pala* or *fanga* but not of *pala+fanga*). The avoidance of redundant information also seems to explain the deletion of *rellotge* in *rellotge despertador* (clock alarm.clock).

Within the same group (64b) we have included colour names, like *blau gris* (blue grey) 'greyish blue', whose status is controversial in the literature. Regarding the syntactic category of each constituent, most authors treat them as the sum of two adjectives (Mascaró 1986, Gavarró 1990b, Scalise 1992, Gràcia 2002[163]). By contrast, we consider that they are the sum of two nouns. Cases of underived adjectival AA compounds are rare (if they exist at all). In the subsection on adjectival compounds, we will see that some AA compounds are based on an underlying AN compound (i.e. the adjective modifies the noun), which undergoes further adjectival suffixation (cf. *mal+humor+at* bad+humour+ed (83c)) and that the AA sequence internal to alleged exocentric CRD compounds is not a compound (e.g. *un diccionari anglès-català* 'an English+Catalan dictionary' (83b)).

Regarding the semantic head, Mascaró (1986) understands that the compound denotes an intermediate shade between the two colours. Gavarró's (1990b: 173-174) position is misleading: on the one hand, it is similar to Mascaró's in that she also understands the compounds as having an intersective reading, but on the other hand she argues that "the first member of the compound may be a modifier of the second one", namely that these compounds are right-headed (e.g. *blau+verd* (blue+green) 'bluish green'). Cabré & Rigau (1986) maintain that they have no head. Mascaró's position can be maintained if, for example, *blau+verd* (blue+green) and *verd+blau* (green+blue) are

---

[163] We find Gràcia's (2002) position contradictory. She takes as a noun the head *verd* of the compound *verd oliva* (green olive) 'olive green', but considers *blau gris* (blue grey) the sum of two adjectives.

interpreted identically; Gavarró's view can hold if *blau+verd* (blue+green) describes a type of green and *verd+blau* (green+blue) a type of blue; and Cabré & Rigau's position can be backed up if the compound does not refer to any type of blue or green. None of the positions discussed above can be corroborated by looking at *blau+verd* (blue+green), because such a word has become lexicalized and the two colours have come to denote a new colour, which is a shade between the two colour names. This is basically Mascaró's (1986) position but note that the opposite order of colours does not refer to the same shade, contrary to his prediction. This result is due to the fact that only *blau+verd* (blue+green) is lexicalized and *verd+blau* (green+blue) is not.

Only non-lexicalized combinations of colour names will be considered. Speakers (ourselves included) agree that the first noun is the semantic head while the second one acts as a modifier, so that *blau gris* (blue grey) is a shade of blue and *gris blau* (grey blue) is a shade of grey. In other words, the compounds are left-headed semantically: they denote a shade of the colour in the first position, which has some properties of the second colour. Concerning the formal head, there is some variation among speakers on where to place number and gender agreement: for some speakers the compound is invariable; for other speakers inflections are placed on the second noun; and still other speakers admit both possibilities. As for the diminutive marking, it tends to be placed on the second noun, which at first sight may seem contradictory: if the first noun is the semantic head, inflections would be expected to occur on it. However, if we look at the semantics more carefully, we see that the compound refers to the colour in first position with some properties of the colour in second position. By placing the diminutive suffix on the second noun, we decrease the properties of the second colour which will be attributed to the first noun and give more prominence to the first colour, which is precisely the interpretation speakers get. We tentatively conclude that compounds made up of two colour names are no different from compounds like *faldilla pantaló* (64a), namely endocentric ATR compounds.

What are claimed to be CRD compounds by a number of authors are exemplified below:

(65)   a. aigua+mel (water+honey) 'a drink made of water and honey', aigua+neu (water+snow) 'sleet', aigua+pedra (water+hailstone) 'a mixture of water and

hailstone', blau+grana (blue+garnet), and sal+pebre (salt+pepper) 'a mixture made of salt and pepper'.

b. **(diccionari) Alcover-Moll** ((dictionary) Alcover+Moll), (vol) Àustria-Hongria ((flight) Austria+Hungary), nord-est (north+east), (míssil) terra-aire ((missile) land+air), (efecte) glaç-desglaç ((effect) frost+defrost), (relació) qualitat-preu ((relation) quality+price), and sud-oest (south+west).

c. alcalde president (mayor president), bomber escalador (firefighter climber), **cuina-menjador** (kitchen+dining.room), diccionari-enciclopèdia (dictionary+encyclopaedia), entrenador jugador (coach player), magistrat jutge (magistrate judge), menjador-sala d'estar (dining.room+sitting.room), poeta pintor (poet painter), and professor investigador (lecturer researcher).

We regard the forms in (65a) not as compounds but as lexicalizations of syntactic coordinate phrases, which are probably perceived as simplex forms. This view can explain why plural marking goes to the end of the word if it is possible: *sal+pebre+s* (salt+pepper-PL). This view can also explain the existence of verbs like *salpebrar* (salt+pepper-IS) with the verbal inflection at the end, formed on the basis of *sal+pebre*. It is difficult to tell which element determines the gender of the word in most cases because the two elements are of the same gender, but *sal+pebre* seems to indicate that it is the second element: *[[sal]ₒₑₘ+[pebre]ₘₐₛ𝒸]ₘₐₛ𝒸*. With the exception of *blau+grana*, the rest of the forms mainly refer to an entity (a mixture) made up of the two elements of the compound (recall Lieber's 2008 'mixture' label for these compounds). In the case of *blau+grana* (blue+garnet), it refers to the Barcelona football club, whose players' uniform consists of blue and garnet stripes, and not of stripes which are of a shade which may result from mixing blue and garnet. Whether the result of coordinating the two elements is an indistinguishable mixture or an entity defined by the two elements separately is an extragrammatical issue.

Gràcia (2002: 819-820) observes that forms like *aigua+mel* (water+honey) 'a drink made of water and honey' (65a) are infrequent in comparison to similar cases which contain the overt conjunction *i* 'and' signalling the coordinate relation between the two elements: *all+i+oli* (garlic+and+oil) 'garlic and olive oil vinaigrette', *col+i+flor* (cabbage+and+flower) 'cauliflower', *all+i+pebre* (garlic+and+pepper)

'sauce made of garlic and pepper', *cap+i+pota* (head+and+foot) 'stew that contains the soft parts of a pork/beef/veal's head and feet', *nap+i+col* (turnip+and+cabbage) 'a round turnip'. Cabré & Rigau (1986) distinguish between complex forms whose semantics is of addition: *plats-i-olles* (plates+and+pots) 'plates and pots shop', *all+i+olli* (garlic+and+oil) 'garlic and olive oil vinaigrette' (which are comparable to the semantics of the forms in (65a)) from those forms which resemble the semantics of ATR compounds in (64a) (e.g. *faldilla pantaló* (skirt trousers) 'skort') like *nap+i+col* (turnip+and+cabbage) 'a round turnip': a turnip which has the shape of a cabbage. We take all these forms with the conjunction *i* not as proper compounds but as lexicalized syntactic phrases, similar to the forms listed in (65a). Some authors (e.g. Cabré & Rigau 1986) call these formations 'synaptic compounds' because they are phrasal in form but behave like simplex words. The hypothesis that the forms in (65a) come from phrases seems to be confirmed by Gràcia (2002: 823), who observes that *blau+grana* initially had the conjunction –*i*-, which is still present in parallel constructions like *blanc-i-blau* (white+and+blue), a combination of colours which identifies another football club, the R.C.D. Espanyol. If such a view is correct, the forms in (65a) cannot be considered (CRD) compounds.

Pérez Saldanya et al. (2004: 260) and B&S (2005), among others, have called the forms in (65b) CRD exocentric compounds. Their argumentation is as follows. The exocentricity of compounds like *nord-est* (north+east) comes from the fact that they refer to an intermediate direction between the two cardinal points denoted by the compound, namely somewhere between north and east. In addition, these compounds do not pluralize and the masculine gender could come from either component (no visible formal head). The CRD relation is not overtly expressed but a conjunction is implicit: e.g. in *nord-est* (north+east) the intermediate direction is somewhere between north *and* east, and in *(relació) qualitat-preu* ((relation) quality+price), the relation is established between quality *and* price. They regard this compound type as productive.

There is no agreement as to how to treat *nord-est* (north+east) and similar forms indicating cardinal points: Cabré (1994) and Pérez Saldanya et al. (2004) treat them like adjectival AA compounds whereas Cabré & Rigau (1986) and Gràcia (2002) treat them like NN compounds. A fact which suggests that it is nominal rather than adjectival is that if *nord-est* (north+east) is placed in a comparative construction, the result sounds odd: #*El tresor està més nord-est* (The treasure is more north+east). An NN analysis is

adopted in the present work. Rainer & Varela (1992) and Val Álvaro (1999) also treat the Spanish counterparts as the sum of two nouns (see Rainer & Varela 1992: 126-127 for the restrictions on these compounds, which also apply to the Catalan counterparts).

Having said that, we want to argue against treating the forms in (65b) as CRD exocentric compounds. We think that what has been called a CRD compound is not a compound at all, but a syntactic phrase which behaves like a simplex word. As such, it can be inserted in the nonhead position of a compound. To illustrate the point, *un vol Àustria-Hongria* (a flight Austria+Hungary) refers to a flight that goes from Austria to Hungary, but *el tractat Àustria-Hongria* (the treaty Austria+Hungary) refers to the treaty which involves both Austria and Hungary. That is, *Àustria-Hongria* by itself does not make much sense unless the implicit conjunction is made explicit and then we have a fully visible syntactic structure, as in *Àustria **i** Hongria* (Austria **and** Hungary); or it is turned into a word and inserted into the nonhead position of a compound, as is the case of the compounds in (65b). In other words, we consider complex words like *un vol Àustria-Hongria* compounds not by virtue of *Àustria-Hongria* (as has been claimed so far in the literature) but by the subordinating relation established between the head *vol* and the non-head *Àustria-Hongria*. As such, they will be considered endocentric SUB compounds: the inherent semantics of the head of the compound, which can be seen in terms of Pustejovsky's (1995) qualia structure, determines the type of relationship between the head and the non-head.[164] Similar findings for the interpretation of compounds are reached by Wisniewski (1996): his study establishes different interpretation strategies used by speakers like 'property mapping', which derives the modificational status of the non-head. In some cases (probably due to their frequent use and some degree of fixation), the role perfomed by the head seems redundant and the non-head alone is used with the same semantics as when it is compounded. That is the case of *nord-est* (north+east), for example.

For some authors (cf. Rainer & Varela 1992, Val Álvaro 1999, B&S 2005), the complex words in (65c) are unquestionably CRD endocentric compounds ('simultaneous' compounds in Lieber's 2008 terms). Recall the discussion of the English counterpart in (36a). Following B&S's (2005) classification, the endocentricity

---

[164] See e.g. Demonte (1999) for an implementation of Pustejovsky's (1995) model to explain the semantics of some combinations in Spanish, one of which is the N+A construction, where the adjective is relational (e.g. *tren eléctrico* 'electric train' vs. *central eléctrica* 'electric power station').

is explained as follows. Semantically, each noun is a hyperonym of the whole compound: *un alcalde-president* (a mayor+president) is both *un alcalde* (a mayor) and *un president* (a president), which also illustrates the coordinate relation between the two elements. Formally, the two nouns tend to be marked for plurality, which is exemplified in *uns alcalde+s-president+s* (some mayor-PL+president-PL), although gender is determined by the first noun: *[un [menjador]$_{masc}$-[sala d'estar]$_{fem}$]$_{masc}$* (a dining.room+sitting.room). On this view, the compounds are double-headed.

Pérez Saldanya et al. (2004: 255) observe that these compounds are not very productive in Catalan, an opinion which contrasts with that of Rainer & Varela (1992: 125-126), according to whom coordinate NN compounds are "highly productive" in Spanish. They mention that although the semantic fields are not restricted, some are more salient than others, such as social roles/jobs, as in *poeta-pintor* (poet+painter) and *cantante autor* (singer author) 'singer-songwriter', and places, as in *panadería-pastelería* (bakery+pastisserie) and *restaurante-centro social* (restaurant+centre.social). Val Álvaro (1999: 4781-4782) illustrates the CRD compounds of Spanish with *entrenador jugador* (coach player), *autor director* (author director) (jobs and activities), *bar restaurante* (bar restaurant), *otoño invierno* (autumn winter) (locations: spatial and temporal), *falda pantalón* (skirt trousers), *pañal braguita* (nappy knickers) (objects), among others. Like in Catalan (65c), both constituents pluralize: *poeta+s-pintor+es* (poet-PL+painter-PL) vs. **poeta+s-pintor* (poet-PL+painter), **poeta-pintor+es* (poet+painter-PL). Villalva (1992: 209, 211) observes the same facts for Portuguese: the plural of *actor-encenador* (actor+producer) is *actor+es-encenador+es* (actor-PL+producer-PL).

Now let us consider why we do not think the complex words under analysis are compounds but (reduced) syntactic structures with a coordinate relation (compare Olsen 2001). As Rainer & Varela (1992: 119) put it, in *clérigo poeta* (priest poet) "both professions are presented as equally constitutive of the **person** referred to (…)". Similarly, in Villalva's (1992: 209) terms, *actor-encenador* (actor+producer) refers to "**someone** who is both an actor and producer". All this shows that if we treated such forms as compounds, they would not be endocentric but exocentric in B&S's (2005) classification: the two elements forming the alleged compound refer to an entity which is located outside it ('person' in Rainer & Varela's definition and 'someone' in Villalva's). This conclusion is similar to that of Levi (1978). However, we want to defend the view that they are not exocentric CRD compounds either.

That we are facing truly coordinate structures can be seen from examples provided by Val Álvaro (1999: 4782, from Rainer 1993: 255), where more than two nouns can stand in a coordinate relation: *director-guionista-creador-productor* (director+scriptwriter+creator+producer) and *su papel de amante-madre-esposa* (her role of lover+mother+wife). What these examples seem to show is asyndetic coordination, namely coordination of several conjuncts with the coordinating conjunction being omitted. This agrees with the known fact that coordination is attested in syntax. Whether coordinate structures of the type discussed here are possible in compounding (or morphology in general) is not yet clear (cf. footnote 139), and we assume that they are not. In short, when the juxtaposition of two (or more) nouns has a coordinate reading we will understand that it is a case of syntactic coordination, and not of compounding. However, a coordinate reading of the forms in (65c) is not the only available one. They can also be treated as endocentric SUB/ATR compounds, similar to *faldilla pantaló* (skirt+trousers) 'skort' in (64a), which is perceived as a type of skirt with some properties characteristic of trousers. For example, *cuina-menjador* (kitchen+dining.room) (65c) can be seen as a kitchen that can also be used as a dining room (endocentric SUB) and *entrenador jugador* (coach player) as a coach who also plays/can also be a player (endocentric SUB/ATR). The potential dual membership of the same compound to SUB and ATR types will be further discussed in section 2.4, and for now they will be treated as SUB compounds in Table 2.4. The interpretation of originally intended CRD structures as endocentric SUB/ATR compounds is facilitated by the fact that CRD structures tend to get fixed in the language in a particular order, which may help develop an interpretation different from the one intended initially (see also Zwanenburg 1992b: 224-226).

The distinction between syntactic coordinate structures (or CRD compounds for some authors like Rainer & Varela 1992, Val Álvaro 1999, and B&S 2005) and endocentric SUB/ATR compounds can be made formally. Formal marking (e.g. plural marker, the diminutive suffix) will be placed on both nouns when they form a coordinate construction, but tends to be on the first noun only when they form an endocentric SUB/ATR compound with the first noun being the head. Gender cannot help distinguish between the two constructions, because it always comes from the first element. However, note that double plural inflection does not uniquely identify a syntactic construction. Recall that there were also some cases of double plural inflection in the endocentric ATR compounds discussed in (64a, b). In addition to plural marking,

other tests need to be applied (e.g. diminutive inflection, hyponymy test). In other words, *faldille+s pantaló* (skirt-PL trouser) is clearly a compound: the plural inflection only attaches to the first noun, the diminutive suffix also applies to the first noun (*faldill+eta pantaló*), and the compound is a hyponym of the entity denoted by the first noun. By contrast, *bomber+s escalador+s* (firefighter-PL climber-PL) could be either a case of asyndetic coordination (people who are both firefighters and climbers) or a compound (firefighters who are also climbers, firefighters who can also work as climbers). The compound status of an NN sequence becomes visible when the diminutive suffix is placed only on the first noun: *bomber+et escalador*. In short, the behaviour of formal markers and the semantics of the NN sequence indicate whether such a sequence is a syntactic or a morphological product.

Taken at face value SUB compounds can be divided into endocentric (66a, b) and exocentric (67a, b) compounds following B&S's (2005) classification. On closer examination, though, exocentric compounds seem to be non-existent, and as for the remaining types, none of them is productive nowadays, which explains why some authors (e.g. Cabré 1994) do not mention them.

(66) a. aigua+batent (water+beat-PRES.PPLE) 'place where water beats', aigua+vessant (water+flow-PRES.PPLE) 'the slope of a mountain where water flows down', casa+tinent (house+own-PRES.PPLE) 'person who has a house in a village and is the head of the family', fir(a)+andant (fair+go-PRES.PPLE) 'a seller who goes from fair to fair', lloc+tinent (place+own-PRES.PPLE) 'lieutenant, deputy', missa+cantant (mass+sing-PRES.PPLE) 'a priest who celebrates mass for the first time', missa+dient (mass+say-PRES.PPLE) 'a priest who celebrates mass for the first time', **terra+tinent** (land+own-PRES.PPLE) 'landowner', and vi(a)+anant (path+go-PRES.PPLE) 'passer-by'.

b. **auto+pista** (car+track) 'motorway', cap-rodo (head+'round') 'vertigo', ferro+carril (iron+rail) 'railway', sal+pàs (salt+step/pass) 'Catholic ceremony in which the priest sprinkles holy water and salt over the gates of the houses', sang+fluix (blood+flow) 'blood flow', and terra+trèmol (earth+quake) 'earthquake'.

There are just a few compounds of the type illustrated in (66a) and most of them are formations dating back to the XIV century (Gràcia 2002: 803). In B&S's view, they are endocentric SUB compounds. As for the grammatical relation between the two constituents of the compound, the non-derived noun (the first noun) is an argument of the deverbal noun (the second noun), hence the label SUB compounds. Recall that argument inheritance is not possible (cf. (34) in this chapter), and consequently the internal structure of such complex forms must have been [[NV]$_V$N]$_N$. Although such structure must have been visible to the speakers initially and might still be apparent for some speakers (e.g. *missa+dient* (mass+say-PRES.PPLE)), we believe there has been a change to [[N[VN]$_N$]$_N$ and ultimately to [NN]$_N$ (eg. *aigua+vessant* (water+flow-PRES.PPLE) 'the slope of a mountain where water flows down'). The fact that some forms are archaic (e.g. *lloc+tinent* (place+own-PRES.PPLE) 'lieutenant, deputy') and that no new formations can be created seems to confirm this hypothesis (cf. Oniga 1988, in Gràcia 2002: 803). The internal word order of the elements in the existing compounds may follow from the fact that they are Latin-based formations, a language with an OV word order (see Pinkster 1990 for discussion on Latin word order). Reinterpreted as NN compounds, the second constituent seems to function as the semantic head generally, although in some cases the form does not exist independently. For example, *un missa+cantant* (mass+sing-PRES.PPLE) would be *un cantant de missa* (a singer of mass), interpreted metaphorically as the priest who says mass for the first time. Formally, the deverbal noun seems to be the head as well: the plural marker goes on the deverbal noun when the compound is pluralized, as in *terra+tinent* (land+owner) – *terra+tinent+s* (land+owner-PL). Given the productive pattern of NN compounds with the first noun as the head (cf. 64a, b), we believe that the compounds under analysis will all come to be reanalysed as simplex words, with some having already undergone such a process, which explains the speakers' inability to create similar forms.

Concerning the compounds in (66b), they are very similar to the compounds just discussed in some respects. They are also right-headed semantically and formally (hence, their endocentricity): an *autopista* (car+track) 'motorway' is a type of *pista* 'track' and the second element bears plural marking and determines the gender of the whole compound. Take for instance, the plural of *auto+pista* is *auto+piste+s* (car+[track-PL]) 'motorway' and the compound is feminine because *pista* is feminine: *[auto]$_{masc}$-[pista]$_{fem}$]$_{fem}$*. The semantic and formal heads coincide in these compounds,

making them endocentric. They are SUB compounds because the first noun is subordinated to the head on the right: an *auto+pista* (car+track) 'motorway' is *una pista d'autos* (a track of+cars) and a *terra+trèmol* (an earth+shake) 'earthquake' is *el tremolor de la terra* (the shake of the earth). In fact, some cases could be assimilated to the previous group in the sense that the second noun is deverbal (e.g. *sal+pàs*). However, such cases have been included in (66b) because of the absence of a present participial form *–nt* and the distant connection with the verb: e.g. *sal+pàs* has acquired a completely new meaning (see above), which makes the link to the verb very remote. This is why the gloss provides nouns for the second element of compounds like *salpàs* and *sangfluix* (although originally they come from past participles).

As for the internal word order of the compounds, the dependent-head order of some compounds can be explained by the fact that they are quite old and may be Latin-based (e.g. *sal+pàs*, *sang+fluix*, *terra+trèmol*), but other compounds are more recent and appealing to a Latin source seems unwarranted. Gràcia (2002: 807, fn. 22) suggests that for the non-Latin-based compounds, the first constituent could be seen as a prefix, a proposal which we do not share since the first element has the same meaning as the independent counterpart (e.g. *ferro* in *ferro+carril* (iron+rail) 'railway'); this suggests that we are dealing with the same lexical item in both syntax and morphology. Note that this view does not prevent the elements of the compound taking on other meanings: *ferro+carril* can refer, in addition to a type of road, to the vehicle which can be seen on such roads. What really seems to explain the internal word order of these words is the fact that they are borrowings from languages with right-headed morphology like English (see Piera & Varela 1999: 4384 for the same findings in Spanish). This, in turn, explains the fact that speakers cannot create new compounds of this type. All in all, the existing compounds are either Latin-based or borrowings from languages with right-headed morphology.

(67)   a. mesclant+aigües (mix-PRES.PPLE+waters) 'place where waters mix', and portant+veus (bring-PRES.PPLE+voices) 'lieutenant'.

       b. aigua+mans (water+hands) 'water which can be used to wash one's hands' or 'a vessel which contains water to wash one's hands', ball+manetes (dance+little hands) expression in child language which means 'to clap one's hands', boca+màniga (mouth+sleeve) 'cuff', cap+vespre (head+evening) 'dusk',

fil+ferro (thread+iron) 'wire', mare+perla (mother+pearl) 'mother-of-pearl', pasta-dents (paste+teeth) 'toothpaste', and vora+via (near+rail/track) 'sidewalk'.

If the two examples illustrated in (67a) could be analysed as compounds, they would be exocentric SUB compounds in B&S's (2005) scheme: the non-derived noun in second position serves as the complement to the deverbal noun in first position (note the nominal suffix –nt which changes the verb into a noun). As can be seen from the paraphrases in (67a), *mesclantaigües* refers to a place where waters mix and *portantveus* refers to a person who speaks in the name of others. These compounds are then exocentric semantically. The formal head is difficult to establish: the second noun is already plural and the compounds are invariable. If the deverbal noun were the formal head, we would expect it to be able to inflect, contrary to reality. What these forms seem to show is a lexicalized syntactic phrase, with the preposition intervening between the two nouns being deleted: e.g. *un mesclant d'aigües* (a mixing of+waters). Note that *portantveus* is an old form, not used nowadays. In short, they cannot be treated as compounds. (For a different view, see Cabré & Rigau (1986), who treat such forms as [VN]$_N$ compounds).

Again, if the complex words in (67b) can be considered compounds, they are also exocentric SUB compounds (according to B&S's 2005 classificatory scheme) and are not productive. Authors like Cabré & Rigau (1986) have erroneously considered them productive because they have included in this group compounds like *gos llop* (dog wolf) 'wolf hound', which we regarded as endocentric ATR (cf. 64a). As for the compounds under analysis, they are left-headed semantically: *fil+ferro* (thread+iron) 'iron' is a type of thread. In these compounds, though, the semantic head does not coincide with the formal head. Although the gender of the compound is determined by the semantic head, plural marking goes to the end. Final plural marking can be explained by the fact that these compounds are lexicalized and are not seen as two separate words put together. As a result, plural marking treats the word as indivisible and is placed at the end of the word: *fil+ferro* (thread+iron) 'wire' – *fil+ferro+s* (thread+iron-PL). In fact, whether the existing forms in (67b) once constituted a compound type is not totally clear to us. *Fil+ferro* seems to have come from the Latin *filum ferreum* or *filum ferri* in the 19thC (GDLC). Other forms like *aigua+mans*, *ball+manetes*, among others, are clearly derived from phrases: e.g. *aigua+mans* dates

back to the 13thC and comes from the expression *donar aigua a mans* (give water to hands). The phrasal source of some forms in (67b) is confirmed by the co-existence of *pasta-dents* and *pasta de dents* (paste of teeth) (cf. Pérez Saldanya et al. 2004). Preposition dropping in phrases seems to be a more common phenomenon in other Romance languages like Italian: *capo stazione* (head station) 'station master', *punto vendita* (point sale) 'point of sale', *capo+sezione* (head+department) 'department head' (Bisetto & Scalise 1999: 34) and Spanish: *tren mercancías* (train goods) 'freight train', *gel ducha y baño* (gel shower and bath) 'shampoo and body gel' (Rainer & Varela 1992: 120) (for other examples in Spanish, see Val Álvaro 1999: 4828, and for some examples in Portuguese, see Villalva 1992). To conclude, it seems that the forms listed in (67b) are not part of an existing compound type. The fact that speakers cannot create new forms based on this pattern confirms our conclusion.

[VN]_N compounds

This type is very productive, and is regarded by Mascaró (1986: 58) as the most productive compound of all.[165] Consequently, unlike the examples in (67b), the absence of functional material between the two compounding components in the forms of (68) (e.g. *enterra+(*els)+morts* (bury+the+dead)) cannot be a sign of the lexicalization of syntactic phrases. In the present case, new forms can be easily created and functional material does not intervene between the two compounding constituents. If they were a syntactic product, functional material would be present but is not, which indicates that these forms are clearly compounds (see also Gràcia 2002). Also, the productivity of this compound type is not hindered by the presence of some non-transparent compounds, like *venta+focs* (blow+fires), whose original meaning of 'an instrument used for stirring up a fire' has been replaced by 'Cinderella'.

As for the semantics, most of these compounds denote agent nouns, e.g. *enterra+morts* (bury+deaths) 'gravedigger' and instrumental nouns, e.g. *porta+avions* (carry+planes) 'aircraft carrier', but they can also refer to locations, as in *guarda+roba* (keep+clothing) 'wardrobe'. Some can also denote animals like *espia+dimonis* (spy+demons) 'dragonfly', and plants like *escanya+llops* (strangle+wolves) 'a poisonous grass'. Sometimes some of these compounds have a negative connotation

---

[165] The same is true of other Romance languages (cf. Duarte & Alsina 1986: 147, Val Álvaro 1999: 4788, Bok-Bennema & Kampers-Manhe 2006, Scholz 2009 in general; and e.g. Varela 1989, 1990, 2005 and Buenafuentes 2007 for Spanish, and Scalise 1992: 191 for Italian).

like *pixa+tinters* (piss+inkwell) 'penpusher', and others can be interpreted metaphorically and be seen as habits, as in *somia+truites* (dream+trouts) 'visionary' (cf. e.g. Cabré & Rigau 1986, Gràcia 2002, Grossmann 1986, Mascaró 1986).

(68)   bufa+forats (blow+holes) 'a type of insect', bufa+focs (blow+fires) 'an instrument used for stirring up a fire', bufa+canyes (blow+canes) 'a musical instrument', bufa+núvols (blow+clouds) 'a conceited person', **busca-raons** (look.for+reasons) 'troublemaker', cobre+llit (cover+bed) 'bedspread', cobre+taula (cover+table) 'table cloth', compta+gotes (count+drops) 'dropper', enterra+morts (bury+deaths) 'gravedigger', escalfa+cadires (heat.up+chairs) 'somebody who outstays his/her welcome', escalfa+llits (heat.up+beds) 'bed warmer', escalfa+panxes (warm+bellies) 'fireplace', escanya+llops (strangle+wolves) 'a poisonous grass', escorre+plats (drain+plates) 'plate rack', escura+butxaques (clean+pockets) 'swindler' and 'slot machine', escura+dents (clean+teeth) 'toothpick', escura+xemeneies (clean+chimneys) 'chimney sweeper', esgarria+cries (lead.astray+offspring) 'wet blanket', espanta-sogres (scare+mothers-in-law) 'party blower', espia+dimonis (spy+demons) 'dragonfly', guarda+boscos (keep+forests) 'forest ranger', guarda+espatlles (keep+backs) 'bodyguard', guarda+joies (keep+jewels) 'jewel case', guarda-roba (keep+clothing) 'wardrobe', llepa+fils (lick+threads) 'to be picky', lliga+cama (tie+leg) 'garter', mata+parents (kill+relatives) 'a kind of mushroom', munta+càrregues (mount+loads) 'lift', neteja+vidres (clean+windows) 'window cleaner', obre+ampolles (open+bottles) 'bottle opener', obre+llaunes (open+tins) 'tin opener', par(a)+aigües (stop+waters) 'umbrella', para+brises (stop+breezes) 'windscreen', para+llamps (stop+lightning) 'lightning conductor', para-xocs (stop+crash) 'bumper', passa+mà (pass+hand) 'handrail', passa+muntanyes (pass+mountains) 'balaclava', passa+port (pass+port) 'passport', passa+temps (pass+time) 'entertainment', pela+canyes (peel+canes) 'to be a nobody', penja-robes (hang+clothes) 'hanger' and 'clothes rack', perdona+vides (forgive+lives) 'a tough person', pica+porta (knock+door) 'door-knocker', pixa+tinters (piss+inkwell) 'penpusher', porta+avions (carry+planes) 'aircraft carrier', porta+equipatge (carry+luggage) '(car) boot', porta+estendard (carry+flag) 'flag holder', porta+veu (carry+voice) 'spokesperon', rasca+cels (scrape+skies)

'skyscraper', rebenta+pisos (break+flats) 'thief', renta+plats (wash+dishes) 'dishwasher', roda+món (travel+world) 'globetrotter', salta+taulells (jump+counter) 'shop assistant', salva+vides (save+lives) 'lifeguard' and 'life jacket', somia+truites (dream+trouts) 'visionary', talla+paper (cut+paper) 'paper cutter', toca+campanes (ring+bells) 'feather-brained person', trenca+closques (break+heads) 'jigsaw puzzle', trenca+nous (break+nuts) 'nut cracker', trenca+colls (break+necks) 'death trap', and venta+focs (blow+fires) 'an instrument used for stirring up a fire' and 'Cinderella'.

Whether the compounds in (68) are endocentric or exocentric is open to debate. There are at least two different views regarding the status of the first element: some authors view it as a verbal form (e.g. Fabra 1956, Cabré & Rigau 1986, Mascaró 1986, Moll 1952, 1975, Gavarró 1990b for Catalan; Contreras 1985 for Spanish; Scalise 1992, Bisetto & Melloni 2008 for Italian, and Bok-Bennema & Kampers-Manhe 2006 and Scholz 2009 for Romance in general)[166] while other authors take it as a deverbal nominal with an agentive value (e.g. Grossmann 1986 for Catalan, and Varela 1989, 1990, 2005, Clements 1992, and Val Álvaro 1999 for Spanish).

Most authors who adopt the first position (i.e. the first element is a verb) treat the compound as exocentric: the compound is a noun and although the second element of the compound is also a noun, the nominal status of the compound cannot come from the second element since neither the formal head (with the exception of the syntactic category and in some cases plural marking) nor the semantic head is located there (but see Fábregas & Scalise 2008 for a different view). The majority of compounds have the second element in the plural, irrespective of whether the compound refers to one entity or to more than one (e.g. *un/dos busca-raons* (one/two look.for+reasons) 'one troublemaker' and 'two troublemakers'),[167] and the gender of the compound may not

---

[166] Those authors who argue for the verbal status of the first element do not always agree on what kind of verbal form it is (i.e. present indicative, imperative, stem). We will not go further into this issue since what interests us is to find out whether the category of the first element is nominal or verbal, and not the specificities within each category.

[167] The singular and plural forms can coexist in some compounds without any change in meaning, as in *para+brisa/para+brises* (stop+breeze/stop+breezes) 'windscreen(s)' (see Lang 1992: 107 for the same phenomenon in Spanish), but in other cases the singular/plural contrast brings with it a change in meaning, as in *pica+porta* (knock+door-SG) 'door-knocker' and *pica+porte+s* (knock+door-PL) 'person who goes from door to door asking for charity' (cf. Grossmann 1986: 158). However, the majority of this compound type has the second element as either singular or plural, which can be predicted to a certain extent: e.g. if the noun denotes a unique entity, it tends to be singular, as in *roda+món* (travel+world) 'globe-trotter' (for other predictions, see Wheeler 1977: 238).

coincide with the gender of the second noun, as in *[cobre+[taula]$_{fem}$]$_{masc}$* (cover+table) 'table cloth'. The majority of compounds are treated as being masculine, except for the compounds which refer to agent nouns, for which there is no gender inflection: *un noi/una noia busca-raons* (a boy/a girl look.for+reaons) 'boy/girl troublemaker'. Semantically, the compound does not denote a kind of the entity denoted by the second element: a *cobre+taula* (cover+table) is not a type of 'table' but rather a piece of clothing spread over it. The semantic head is located outside the compound. These paraphrases also make it clear that the noun is the internal argument of the verb; hence SUB compounds (on which more will be said below).

Contrasting with this view, Bisetto & Melloni (2008) maintain for Italian that the first element is verbal, but that the compound is endocentric. On their analysis, the verb and the noun get together in syntax and then the outcome becomes a noun by means of a nominalizing zero suffix. Although we agree that there is a nominalizing affix responsible for the nominal nature of the compound, we do not share the view according to which the merger of the verb and the noun is syntactic. We will not dwell on their analysis, but just note that on this account there is no reason why the external argument is never present.[168] In addition, to account for the semantics of the compound, there is a GenericAspectPhrase (GAP) above VP, which hosts the thematic vowel responsible for the generic reading of the construction, so they claim. The role of such GAP is not clear to us: all roots have to go through GAP to get the thematic vowel and the right semantics. In other words, the movement seems unmotivated since there are no cases in which there is no movement and others in which there is, and the absence of inflection on the verb is not explained either, since verbs inflect for tense and agreement in syntax in order to be well-formed. (See Contreras 1985 and Fábregas & Scalise 2008 for a similar analysis for Spanish and Italian compounds in the sense that the verb and the noun are also merged in syntax). An alternative approach is to merge the verbal stem (root+thematic vowel) with the noun in morphology. If the merger is morphological, then the absence of an external argument follows since there is no position in the tree dedicated to it (but see A&N 2004: 39-42 for a proposal in which the external argument can be present in a compound given a specific context). The presence of a stem, and not

---

[168] One could argue that the verb's external argument role is bound by the R-role of the nominal affix, which would explain why the external theta-role of the verb is not available for assignment to an argument. While this analysis cannot be applied to Bisetto & Melloni's (2008) representation of such compounds, according to which the nominalizing affix is higher than the VoiceP which introduces the Agent, it can be applied to our analysis, according to which the nominalizing suffix comes right after the verb and the noun have merged together in morphology.

of a fully inflected verb, is explained by the fact there are no functional categories like tense in the morphological component. The absence of inflection on the stem explains, in turn, the habitual reading of the compound. (See Bok-Bennema & Kampers-Manhe 2006 for a morphological analysis according to which morphological constructs may be phrasal, as is the case of VN compounds on their view).

If one adopts the second position (i.e. the first element is a deverbal nominal), then one can try to argue that the compound is endocentric. Semantically one could maintain that the compound is a type of entity denoted by the first element which has an agentive value, and that the masculine gender of most compounds is the unmarked gender. Varela (1989: 406-407, 1990: 76) supports this view and argues that the thematic vowel nominalizes the verbal root and is the element endowed with agentivity, or the feature 'actor' which includes both agents and instruments in Varela's view. She uses as an argument for the nominal status of the first element the fact that one can use the first constituent as an agent noun: *un limpia* (a clean) 'a bootblack' would mean the same as *un limpia+botas* (a clean+boots).[169] However, the first part of such a compound - which may be derived by clipping the compound *un limpia+botas* (a clean+boots) 'a bootblack' - may have acquired its nominal and agent-like properties by another means, for example by having a head outside *limpia+botas*. Notice, in addition, that not all compounds can be clipped in the way *un limpia* (a clean) can: one cannot use *un para* (a stop) instead of *un par(a)+aigües* (a stop+waters) 'an umbrella' to refer to the same entity or to any sensible entity. Although the exact restrictions are difficult to pinpoint, clipping seems to be severely restricted and not all first constituents can be used as agent nouns. Furthermore, forms like *un ocupa* (an occupy) 'a squatter', *un limpia* (a clean) 'a cleaner' and *un busca* (a look.for) 'a small radio receiver that beeps, vibrates, or flashes to alert the user to an incoming message which is usually displayed on a small screen' (definition from the MWOD) are more general terms than *un ocupapisos* (an occupy+flats), *un limpiabotas* (a clean+boots) and *un buscapersonas* (a look.for+people), the forms from which the clipped forms are allegedly derived. This view has the unwelcome consequence that clipped forms cannot uniquely identify the compounds, because for example, *un ocupa* 'a squatter' is not necessarily a person who

---

[169] Note that this possibility is not available to Catalan: e.g. *un neteja+vidres* (clean+glasses) 'windowcleaner' vs. **un neteja*. Both *neteja* and *vidre* can turn into agent nouns by means of a suffix: *neteja+dor* (clean+er) and *vidri+aire /vidri+er* (glass+er) 'glassworker'. See Grossmann (1986: 165-166) for examples of this type.

lives in a flat illegally, but they can also live in a house, bungalow, or mansion illegally and may as well be called *un ocupa* 'a squatter'. In addition, if the first element were the real head in the compound, we would expect the first element to inflect for plurality. For instance, we would expect to be able to contrast *un busca-raons* (a look.for+reasons) 'a troublemaker' with *uns busques-raons* (some look.for-PL+reasons) 'some troublemakers', but this is clearly not the case. By contrast, when the compound makes the singular-plural distinction, plural marking goes at the end: *passa+port* (pass+port) 'passport' vs. *passa+port+s* (pass+port-PL) 'passports'.

Here we will just briefly review some other arguments Varela (1989, 1990) provides to favour the left-headed NN structure for such compounds and will see that a better alternative is available. If the thematic vowel is the element responsible for the agentivity of the compound, then one would not expect a suffix with the same role, i.e. with agent-like properties. Cases like Cat. *para+caigud+ista* (para+chute+ist) and *par(a)+aigü(a)+er* (stop+water+er) 'person whose job is to make, mend or sell umbrellas' with the alleged presence of two agent-like affixes seem redundant and cannot be explained. The feature 'agentivity' or 'actor', in fact, may not be the right feature since agentivity is not always involved in the compound, which generally denotes agents and instruments, but can also refer to places, plants, and animals. A feature which seems to better characterize all these denotations is the feature [+c], 'cause change', of Reinhart's (2000, 2001) theta system. (We will not expand on Reinhart's system since it would take us too far afield, but the reader is referred to the original works for details). According to Varela, the non-derived noun in second position is the internal argument of the underlying verb in first position. On this view, the internal argument must be inherited since the verb merges first with the nominalizing thematic vowel before it merges with the internal argument. However, this move is prohibited. Recall, from the subsection of nominal compounds in English (subsection 2.3.1.1), that the idiomatic reading of compounds like *troublemaker* (34) can only be preserved if there is a direct merge of the verb with its internal argument.

Rainer & Varela (1992: 127-130) provide other evidence disfavouring the treatment of superficial V-N compounds as left-headed NN compounds. First, the VN compounds under study and proper left-headed NN compounds have different accentual patterns. Second, while the noun in VN compounds is typically interpreted as an argument of the first element, such interpretation is not available to NN compounds. Third, plural formation is different in VN and left-headed NN compounds: the plural

marker is placed at the end of the word in VN compounds, as in *passa+port+s* (pass+port-PL) 'passports' and on the first noun in NN compounds, as in *camion+s cisterna* (lorry-PL tank) 'tanker lorries'.

In short, it seems preferable to treat such compounds as V+N compounds, with the V being the stem (root+thematic vowel), which is in agreement with the restrictions on this compound type, as noted by Mascaró (1986: 60-61). The head can be an affix, which we take as a suffix, outside the visible constituents of the compound. Semantically, the suffix is endowed with an agentive value, or rather, with a [+c] feature (one could also view that the suffix's R-role binds the verb's external argument role, cf. footnote 168). The suffix is also the formal head: it provides the compound with the nominal category, the unmarked gender in the sense that there is no inflection marking it (i.e. it is identical to the masculine gender), and bears the plural inflection when the compound is plural. The plural marker is only visible in the case the internal argument is singular, e.g. *pica+porta* vs. *pica+porte+s* (knock+door vs. knock+door-PL) 'door-knocker vs. door-knockers'. When the internal argument is plural, the plural markers (the –*s* on the internal argument and the –*s* on the zero suffix) are fused into one and it looks like there is no plural inflection marking the compound as a whole.

As for the type of grammatical relation holding between the two visible constituents, and between these and the [+c] suffix, it is of a SUB type: the noun is the internal argument of the verb, and the NV together is subordinated to the [+c] suffix. To be more precise, the internal argument can only be an affected theme (cf. Rizzi's 1986 distinction between theme 1, which is affected, and theme 2, which is unaffected by the action of the verb). The verb is a transitive action verb unless it is part of a compound which is idiosyncratic in some sense, like the verb *cagar* 'to shit' in *caga+ferro* (shit+iron) 'clinker', *caga+niu* (shit+nest) 'the youngest son/daughter', and *caga+tió* (shit+log) 'Christmas log'. In all these cases the verb is used intransitively, and the last compound is interpreted as a command (for the phonological restrictions this compound type is subject to, see Mascaró 1986: 63).

[PN]$_N$ compounds[170]

According to Mascaró (1986: 64), all PN compounds are exocentric, but as will be seen below, nominal compounds of the PN type can also be endocentric (see below). Like in

---

[170] Recall that not all authors treat such forms as compounds, e.g. Cabré (1994), Cabré & Rigau (1986), Gavarró (1990b) and Turon (1999) treat them as prefixed words.

the survey of English compounding, we will consider as cases of prefixation those complex forms made up of a P+N in which the meaning of the P can no longer be associated with the meaning of the independent counterpart, as is the case of *entre+claror* (between+light) 'a faint light', *sobre+dosi* (over+dose) 'an overdose', and *ultra+dreta* (extreme+right) 'right-wing extremist'.

The compounds have been divided into exocentric (69a) and endocentric (69b). (For more PN compounds, see Fabra 1956).

(69) a. contra+almirall (counter+admiral) 'an officer ranked below a vice admiral', contra+blocatge (counter+blockade), contra+guerrilla (counter+guerrilla) 'a military group that fights against a guerrilla band', contra+verí (counter+poison) 'antidote', **entre+acte** (between+act) 'interval, period of time between acts', entre+cella (between+eyebrow) 'space between the eyebrows', entre+costella (between+rib) 'space between one's ribs', entre+cuix (between+top.leg) 'crotch', sobre+cella (over+eyebrow) 'part above the eyebrows', sobre+taula (over+table) 'table cloth', sota+barba (under+chin) 'space between one's neck and chin, i.e. double chin', sota+cor (under+choir) 'the part of a church under which the singers are placed', sota+escala (under+staircase) 'space under the stairs', sota+teulada (under+roof) 'part of a house just under the roof', sots+governador (sub+governor) 'official ranked below the governor', sots+secretari (sub+secretary) 'officer ranked below the secretary', ultra+mar (beyond+sea) 'overseas', and vice+president (vice+president).

b. avant+cambra (anterior+room) 'anteroom', contra+acusació (counter+accusation), contra+atac (counter+attack) 'counterattack', contra+declaració (counter+statement), contra+exemple (counter+example), contra+oferta (counter+bid), contra+ordre (counter+order), contra+projecte (counter+project) 'project against another project', contra+proposta (counter+proposal), sobre+bena (upper+bandage) 'upper bandage', **sobre+impressió** (over+printing), sobre+jutge (upper+judge) 'superior judge', sota+mola (below+millstone) 'a lower millstone', and sots+director (sub+director).

Following B&S's (2005) classification, the compounds in (69a) would be exocentric SUB: they refer to an entity outside the compound (exocentricity), which is characterized by the compound, and the N in second position is understood as the complement of the P. For instance, *contra+verí* (counter+poison) 'antidote' is a substance that is used to fight against poison and *sota+teulada* (under+roof) is the part of a house located just under the roof. Plural marking goes to the end of the compound, which we understand as being on a head outside the PN compound: *contra+verí* vs. *[contra+veri(n)]s* (counter+poison vs. [counter+poison]-PL). Also, the gender of the compound is not determined by the noun in the compound: e.g. *[sota+[teulada]fem]masc* (under+roof). Similar compounds are observed in other Romance languages, like *sotto+scala* (under+staircase) 'space beneath the staircase' in Italian (cf. Scalise 1992: 193) and *avant-guerre* (before+war) 'pre-war period' in French (cf. Zwanenburg 1992b: 229).

That these compounds are exocentric, as has been traditionally claimed, is not so obvious. If the plural marker goes on a head outside the visible PN compound, the head in question may also be responsible for the semantics, the nominal category of the compound and be endowed with unmarked gender, i.e. masculine. (See the [VN]N compound above which has received a similar treatment, and Gràcia & Azkarate 2000 for the postulation of a similar head outside some prefixed nouns in Catalan like *pre-història* pre+history). This analysis will be adopted in the present work and will prove to be more in line with the overall way compounding works.

Some of the compounds above have a locative reading, the location being either in space or in time: *entre+cuix* (between+top.leg) 'crotch' and *entre+acte* (between+act) 'interval, period of time between acts'. The locative reading also encompasses those compounds which place an official in a hierarchy, as in *sots+secretari* (sub+secretary) 'officer ranked below the secretary' (but see Turon 1999 who treats the semantics of P+N words denoting jobs like *sots+secretari* as derived from the semantics of gradation). Other compounds involve an opposition meaning: *contra+verí* (counter+poison) 'antidote'. The semantics of the compounds fits well with the [-m] feature of Reinhart's (2000, 2001) system, 'm' standing for 'mental state of the participant'. Such a feature may be encoded in the head outside the PN compound.

The compounds in (69b) also seem endocentric SUB, but their internal structure is simpler: there is no head outside PN. The N is the head semantically and formally.

The compounds refer to the entity denoted by the second element of the compound, which is modified by the first element (SUB relation). A *sobre+impressió* (over+printing) is a type of *impressió* (printing), which is characterized by being above already printed material. In most cases the compound denotes a subtype of the entity embodied by the noun which happens to be modified by the preposition whose object is generally of the same nature as the entity denoted by the compound. For example, *contra+atac* (counter+attack) is an attack in response to (against) the enemy's previous attack. Surprisingly, very similar compounds like *sots+secretari* (sub+secretary) and *sots+director* (sub+director) are not placed on the same list: while *sots+secretari* is not a type of secretary but an officer ranked below the secretary, *sots+director* is a type of director, hence their placement in (69a) and (69b) respectively. Regarding the formal head, gender is determined by the second element, and plural marking is placed on the second element as well (e.g. [*contra*+[exemple+s]] counter+example.MASC-PL), thus giving the endocentric feature to the compound. Similar compounds are also present in other Romance languages: e.g. *sotto+commissione* 'subcommittee' and *sotto+bibliotecario* 'assistant librarian' in Italian (although they are regarded as cases of prefixation by Scalise 1992: 193 since in his view (productive) compounding in Italian is all left-headed, which leads him to conclude that such examples cannot be compounds but must be prefixed words) and *avant-projet* (before+project) 'preliminary plan' in French (Zwanenburg 1992b: 229).

[VV]$_N$ formations

The forms in (70) are not productive at all. If they were compounds, they would be treated as exocentric CRD compounds in B&S's (2005) scheme and would be 'simultaneous' in Lieber's (2008) terms. The relation between the two verbs put together is one of conjunction (coordination) and they refer to an entity which is not included in either constituent of the compound (exocentric compounds). In addition, the nominal category of the compound does not come from either constituent of the compound. As for the gender, for example, *alça+prem* (lift+press) is masculine and *suca-mulla* (dip+wet) is feminine (maybe due to the final vowel –*a*, as Gràcia 2002 suggests).

(70)   alça+prem (lift+press) 'lever', cia+voga (row.backwards+row) 'the result or
         effect of turning a boat around by rowing', gira+volta (turn+rotate) 'rotation',

suca-mulla (dip+wet) 'the action of dipping biscuits, bread into one's drink, generally milk or wine'.

Some of these would-be compounds seem to be derived from verbs: *ciavogar$_V$* < *ciavoga$_N$* and *giravoltar$_V$* < *giravolta$_N$*, which makes the nominal forms comparable to the nominal *slam-dunk* and *strip-search* in (41), which also come from verbs.

   We will maintain that the forms with a coordinate relation between the two constituents (70) are lexicalized syntactic phrases with a deleted conjunction, with the result that nowadays they are considered simplex words by speakers. If the coordinate relation were still visible, plural markers would be expected on the two constituents, contrary to reality: *dos *[alce+s+prem+s]* (two *[lift-PL+press-PL]) vs. *dos [alçaprem]+s* (two [liftpress]-PL). In addition, there exist parallel examples to those in (70) but with an overt *i* 'and' in between the two constituents: *estira-i-arronsa* (stretch+and+shrink) 'bargaining', *pèrdues-i-guanys* (losses+and+benefits), *puja-i-baixa* (go.up+and+go.down) '(repeated) action of going up and down', and *va+i+vé* (go+and+come) 'swinging'. These forms will not be considered compounds because they follow the laws of syntax, i.e. they are phrases with the peculiarity that they have become lexicalized and now act as syntactic atoms. From this, we conclude that forms like *alça+prem* (lift+press) 'lever' can be put together with the forms that include an intervening *–i-* (cf. recall that the same phenomenon was found in (65a) where some phrases became lexicalized with the conjunction *–i-* and others without). Note that new forms based on the [VV]$_N$ pattern cannot be created nor can they be assimilated to the *freeze-dry* (49) compound type in English, where the first verb acts as a modifier of the second verb, the head.

   Other Romance languages have similar forms, with the salient feature of being non-productive, like the examples from Catalan. Consider the Italian examples *andir+i+vieni* (go+and+come) 'comings and goings' (Scalise 1992: 177) and *bagna+asciuga* (soak+dry) 'shore' (Fábregas & Scalise 2008), and the Spanish example *sub+i+baja* (climb+descend) 'swinging' (for other forms in Spanish, see e.g. Lang 1992: 104).

[NA]$_N$ formations

There are at least two types of forms that conform to the [NA]$_N$ structure, which are illustrated in (71a) and (71b). Neither can be considered compounds. The former will be

argued to be a syntactic product, i.e. a phrase, while the latter will be a nominalization of an adjectival compound.

(71)   a. correu aeri (mail air) 'mail by air', estrella polar (star polar) 'Pole star', estructura molecular (structure molecular) 'molecular structure', manuscrits suecs (manuscripts Swedish), molí fariner (mill flour) 'flour mill', and tenda reial (tent king) 'king's tent'.

        b. cama+llarg (leg+long) 'wading bird', cap+gròs (head+big) 'tadpole' and 'a figure with a huge head' (typically found in festivals and carnivals), cul+gròs (bottom+big) 'a mushroom: *Amanita ovoidea*', pell-roja (skin+red) 'a red skin, an American Indian', and pit-roig (chest+red) 'a robin'.

As for the forms in (71a), they have been called 'syntagmatic' compounds by Cabré & Rigau (1986), and if we treated them as compounds, they would probably be endocentric SUB compounds (cf. Marchis 2009). Let us consider *estrella polar* (star pole): the underlying non-head noun *pol* 'pole' (in the adjective *polar*) gives the location of the head noun *estrella* 'star' (SUB relation). The gender of the compound is determined by the head (feminine) and plural marking is placed on both constituents, hence an endocentric compound.

However, when subjected to scrutiny, one can observe that these forms are syntactic and that the facts observed above (e.g. having plural inflection on both constituents) follow from the syntactic nature of the phrase. There are a number of tests that compounds pass and the forms in (71a) fail, thus revealing their syntactic nature (cf. subsection 2.3.1; see also Cabré & Rigau 1986, Pérez Saldanya et al. 2004). First, if they were compounds, neither constituent would be able to be modified externally, as in **un [escura[dents]$_{fem}$]$_{masc}$ [brutes]$_{fem}$* (a [clean[teeth]$_{fem}$]$_{masc}$ [dirty]$_{fem}$), but the forms in (71a) show the opposite behaviour: e.g. *els manuscrits suecs* (the manuscripts Swedish) cannot be a compound since *manuscrits* alone can be modified, as shown by *[els manuscrits [suecs] [de Leonardo da Vinci]]*, where *suecs* refers to the manuscripts' owner and *de Leonardo da Vinci* to the manuscripts' author (example from Picallo 2002: 1672). Second, none of the compounding elements can be picked up anaphorically, as in **Vull un escura[dents]$_i$ perquè les$_i$ tinc brutes* (I.want a clean[teeth]$_i$ because them$_i$ I.have dirty) 'I want a toothpick because I have them dirty' but the forms

in (71a) can: *Ja han posat la tenda$_i$ reial al campament base però la$_i$ dels bisbes encara no* (They have put the tent$_i$ royal in the camp base but the$_i$ of.the bishops yet not) 'They have set up the royal tent in the base camp but the bishops' tent hasn't been set up yet'. All things considered, we can conclude that the forms in (71a) are phrases (some may even be seen as collocations due to their recurrent combination, see Lorente 2002), from which it follows that gender is determined by the noun and plural marking is placed on both the noun and the adjective. Given this picture, one would expect the adjective to be able to be modified by a degree quantifier like *molt* 'very' but the result is ungrammatical, *una tenda (*molt) reial* (a tent very royal). The ungrammaticality is explained by the fact that the adjective is relational and they do not accept any kind of degree quantification, since they do not denote a property but a relation (Picallo 2002: 1667-1668).

The forms in (71b) denote an entity which is prototypically characterized by the two members that make up the word: un *cap+gròs* (head+big) is either an animal or a figure which has a big head. Formally, the plural marking is placed at the end and as for the gender, it is the unmarked gender, i.e. masculine. All the forms listed in (71b) can be seen as nominal recategorizations and lexicalizations of transparent adjectival NA compounds (see subsection 2.3.2.3 where more is said about the restrictions to which adjectival compounds of the NA type are subjected). However, newly-created forms based on the [NA]$_N$ pattern can be easily created: *[un panxa+gròs]$_N$* ([a-$_{masc}$ belly-$_{fem}$+big-$_{masc}$]$_N$), *[un cara+vermell]$_N$* ([a-$_{masc}$ face-$_{fem}$+red-$_{masc}$]), and *[un coll-llarg]$_N$* ([a-$_{masc}$ neck-$_{masc}$-long-$_{masc}$]), but notice that they necessarily seem to be derived from an adjectival NA compound. The agreement markers on the NA form can only be explained if such a form is an adjective that agrees in number and gender with the determiner and a null nominal head. That is, a nominal head seems to be missing: the determiner plus the null nominal head determine the gender and number of the adjectival NA complex form before the complex adjective can be used as a noun. Consider the following examples: *[una-$_{fem.sg}$ Ø-$_{N.fem.sg}$ [panxa-$_{fem.sg}$+grossa-$_{fem.sg}$]]$_{N.fem.sg}$* ([a-$_{fem.sg}$ Ø-$_{N.fem.sg}$ [belly-$_{fem.sg}$+big-$_{fem.sg}$]]$_{N.fem.sg}$), *[uns-$_{masc.pl}$ Ø-$_{N.masc.pl}$ [panxa-$_{fem.sg}$ +grossos-$_{masc.pl}$]]$_{N.masc.pl}$* ([a-$_{masc.pl}$ Ø-$_{N.masc.pl}$ [belly-$_{fem.sg}$ +big-$_{masc.pl}$]]$_{N.masc.pl}$), *[unes-$_{fem.pl}$ Ø-$_{N.fem.pl}$ [panxa-$_{fem.sg}$+grosses-$_{fem.pl}$]]$_{N.fem.pl}$* ([a-$_{fem.pl}$ Ø-$_{N.fem.pl}$ [belly-$_{fem.sg}$+big-$_{fem.pl}$]]$_{N.fem.pl}$). The conclusion so far seems to be that the [NA]$_N$ compound type is

unavailable to the language (but see subsection 2.3.2.3 where a special type of NA compound will be shown to exist).

Regarding potential [NA]$_N$ compounds in other Romance languages, the conclusion seems to be the same as that for Catalan. For example, Val Álvaro (1999) argues that forms like *campo+santo* (field+holy) 'cemetery', also spelled *campo santo*, in Spanish are the result of a lexicalized syntactic phrase. As such, we do not consider them compounds. Scalise (1992: 177) also gives the same example in Italian as representative of NA compounds and observes that they are not productive.

[NV]$_N$ formations

Compounds conforming to the [NV]$_N$ pattern seem non-existent. Two forms are listed in (72): it is not clear whether *ben* is a noun functioning as the object of the verb or an adverb modifying the verb (see Duarte & Alsina 1986: 151 for similar forms where the first element can be considered an adverb, as in *ben+viure* (well+live), although they are cases of lexicalization and not proper compounds). Both forms are invariable and the semantic paraphrases suggest that they have undergone a process of lexicalization, which explains the nominal status of the complex word, which does not come from either constituent. The non-existence of an [NV]$_N$ compound type is confirmed by the ungrammaticality of new forms: **sopar+fer* (dinner+cook), **dinar+voler* (lunch+want), and **taula+tenir* (table+have).

(72)   ben+fer (good+do) 'one's rectitude in their actions', and ben+voler (good+want) 'one's good will/love towards others'.

[PA]$_N$ formations

We have found only two examples that conform to the [PA]$_N$ structure, which indicates the non-productivity, and non-existence, of such a type of compound. In addition, both forms denote the same entity. If they could be treated as [PA]$_N$ compounds, they would be exocentric SUB: they refer to a person who has signed at the bottom of a document, with the nominal feature coming from neither constituent. These forms can be seen as nominalizations of adjectival forms (see subsection of adjectival [PA]$_A$ compounds).

(73)   sota+escrit (under+write-PPLE) 'undersigned' and sota+signat (under+sign-PPLE) 'undersigned'.

What follows are different types of complex forms which some authors have considered proper compounds (e.g. Kornfeld 2003 for Spanish, which has the exact counterparts to the Catalan forms listed below), others have called synaptic/syntagmatic compounds (e.g. Benveniste 1974, Cabré & Rigau 1986, Lang 1992, Mascaró 1986, Val Álvaro 1999, Buenafuentes 2007), and still others have omitted from their compounding classification (e.g. Gràcia 2002).

(74)   a. cent+peus (a.hundred+feet) 'centipede', curt+circuit (short+circuit) 'short circuit', mala lluna (bad moon) 'bad mood', mal(a)+ànima (bad+soul) 'a cruel, heartless person', mal+astre (bad+star) 'misfortune', mal+nom (bad+name) 'nickname', mal+son (bad+sleep) 'nightmare', mig+dia (half+day) 'midday', mil+fulles (a.thousand+leaves) 'millefeuille', mil+homes (a.thousand+men) 'a person who boasts about being brave', mitja+nit (half+night) 'midnight', poca pena (little sorrow) 'miserable person', poca solta (little reason) 'thoughtless', sant crist (holy christ) 'crucifix', and tres+peus (three+feet) 'trivet'.

b. aigu(a)+ardent (water+burn-PRES.PPLE) 'liquour, brandy', cel+obert (sky+open-PPLE) 'patio', and sostre+mort (ceiling+die-PPLE) 'loft'.

The forms in (74a) are made up of an adjective or a quantifier and a noun. They are phrasal forms which have become syntactic atoms in the sense that they function as simplex words. They have become lexicalized, i.e. more opaque semantically, sometimes adopting a metaphorical interpretation, as in *mil+homes*. Gavarró (1990b: 172) also treats forms like *poca pena* (little sorrow) 'miserable person' as X'-expressions, namely as part of syntax.

Most of the nouns like those in (74b) have developed out of nominal phrases (a noun and a modifying participle), which have lost their semantic transparency and have become lexicalized as nouns. I disagree with Cabré's (1994) treatment of such compounds as adjectival compounds on a par with *cama+llarg* (leg+long) 'long-legged' and 'wading bird'. Both as phrases and as lexicalized forms, the forms in (74b) are always nominal, unlike *cama+llarg* which can be adjectival (when transparent) and nominal (when lexicalized). Mascaró (1986: 72) includes in this group (74b) forms like

pell+roja (skin+red) 'red-skin', which he separates from compounds like *pit+negre* (chest+black) 'turnstone/a type of bird'. We see no reason to separate them: both are N+A adjectival compounds which have become lexicalised as nouns.

Other clearly lexicalized phrases are those composed of N de N (N of N) and commands: *cul de sac* (bottom of bag) 'cul-de-sac', *ull de poll* (eye of chick) 'callus', and *pa de pessic* (bread of pinch) 'spongecake'; and *no-m'oblidis* (no+me+forget) 'forget-me-not' (see Lorente 2002: 870-872 for other lexicalised sentences, which are used as simplex nouns nowadays). Cabré (1994) accepts that N de N (N of N) are lexicalized syntactic phrases but she treats them as compounds. Such treatment may arise from the traditional view that morphology contains unproductive and idiosyncratic information while syntax is the component where productive and transparent processes take place. If morphology is a generative component on a par with syntax, lexicalized syntactic phrases cannot be placed in morphology. In addition, this view would imply that morphology makes use of the same combining principles available to syntax and that a duplication of rules would be unavoidable, a position not defended here because, for example, the use of PPs seems to be restricted to syntax only. In short, although morphology and syntax share some vocabulary (e.g. lexical categories), they also have their distinctive combinatorial features.

From a syntactic point of view, Kornfeld (2003) also considers compounds the Spanish counterparts of N de N (N of N) (e.g. *ojo de buey* (eye of bull) 'porthole') as well as AN phrases like the Spanish *libre mercado* (free market) 'free market'. Her view of compounding leads her to conclude that they are compounds because such phrases have become syntactic atoms, a fact that we take to mean rather the opposite. If syntactic phrases have become syntactic atoms, it simply means that they behave like simplex words but not that they are compounds. For this reason, we do not treat such forms as compounds. A reason for calling such lexicalized syntactic phrases compounds may be due to their naming function, which is typically associated with compounds (cf. Downing 1977). However, we assume that the naming function may as well arise from the lexicalized nature of such syntactic expressions, and not from the fact that they are compounds. Note that this reasoning also applies to whole sentences (e,g, Sp. *el hágalo-usted-mismo* 'do-it-yourself') and to affixes (Cat. *els pros* 'the pros').

As already noted, some authors have called the forms in (74) synaptic/syntagmatic compounds, as well as those forms discussed in the two previous paragraphs, i.e. N of N structures and sentences used as simplex words. Synaptic/syntagmatic compounds are two different terms which refer to the same entity. Benveniste (1974: 172-173) provides a characterization of *sinapsia* which has been widely adopted since then. Among its main features is the syntactic nature of the construction, the fixed order of its constituents, the lack of a determiner in front of the second constituent, and the presence of a unique and constant meaning. In short, synaptic/syntagmatic compounds are the result of the lexicalization of a syntactic merge that with time has become fixed and adopted an idiosyncratic meaning (Pérez Saldanya et al. 2004: 248).

Of a similar opinion is Buenafuentes (2007), who divides Spanish compounds into three types: syntagmatic (*composición sintagmática*), learned (*composición culta*)[171] and lexical (*composición léxica*). She defines syntagmatic compounds as lexical items which are based on frozen phrases; they are phrasal in form and consequently their members are not spelt as one word (p. 91) (see also Val Álvaro 1999: 4760). Among their defining properties is their frozen nature (i.e. phrases becoming fixed, with the consequence that their order is also fixed), their lack of syntactic autonomy and their semantic unity. Accordingly, she treats forms like N of N as syntagmatic compounds: *ojo de buey* (eye of bull) 'porthole', NA: *caja fuerte* (box strong) 'safe', AN: *media pensión* 'half board' and NN: *pez martillo* (fish hammer) 'hammer fish'. Again, for us, they are not compounds although we also understand them as lexicalized syntactic phrases, with the exception of NN forms, which we believe are formed in the morphological component (see the beginning of the present subsection).

Regarding lexical compounds, they are defined by Buenafuentes as the result of putting two native bases together in terms of morphology, semantics and writing (i.e. spelt as one word). We did not have as a requirement for compoundhood the fact that the compound be spelt as one word: the two compounding elements can also be joined by a hyphen or be two separate words. Buenafuentes' view of compounding leads her to consider as lexical compounds (although of a special sort) forms like *haz+me+(r)reír*

---

[171] Buenafuentes's treatment of learned compounds (neo-classical compounds in our terms) is similar to ours although we do not distinguish, for example between those compounds whose two elements are both from the same classical language (both Greek or both Latin) and those compounds whose elements are each from a different classical language (one from Greek and the other from Latin or viceversa).

(make+me+laugh) 'laughing stock' and *sabe+lo+todo* (know+it+all) 'a know-it-all' (see Val Álvaro 1999: 4838 for other similar forms). They are lexicalized syntactic phrases, which due to their idiosyncratic meaning have come to be spelt as one word. Being spelt as one word is no reason for us to take them to be compounds. That the division between syntagmatic and lexical compounds in Buenafuentes' classification is not neat is clear from some forms which can be spelt as one word or as two words: *noche buena* vs. *nochebuena* (night+good) 'Christmas Eve' (see Lang 1992: 102 for other examples). Furthermore, following Buenafuentes' definition of lexical compounds, learned compounds should also be included within lexical compounds since most of them are spelt as one word. In short, the boundaries of the initial tripartite classification are not clear-cut, which is a problem commonly found in diachronic studies.

This concludes our survey of nominal compounds in Catalan. The results are shown in the following table: one example of each type is given. The forms we have considered as lexicalized phrases are not included, and the compound types whose productivity is null or very limited are included within parentheses.

Table 2.4: Nominal Compounds in Catalan

	NOMINAL COMPOUNDS					
	SUBORDINATE		ATTRIBUTIVE		COORDINATE	
	endocentric	exocentric	endocentric	exocentric	endocentric	exocentric
[NN]$_N$	diccionari Alcover-Moll (65b) (dictionary A-M) cuina-menjador (65c) (kitchen-dining.room) (terra+tinent) (66a) (land+owner) (auto+pista) (66b) (car+track)		faldilla pantaló (64a) (skirt+trouser) 'skort' verd oliva (64a) (green olive) blau+gris (64b) (blue+grey) 'greyish blue'			
[VN]$_N$	busca-raons (68) (look.for+reasons)					
[PN]$_N$	entre+acte (69a) (between+act) sobre+impressió (69b) (over+printing)					
[NA]$_N$	relacions catalano-occitanes (83b) (relations Catalan+Occitan)					

2.3.2.2 Verbal compounds

Four types of verbal compounds can be distinguished: NV, AV, AdvV and PV, although the AV type is argued to have null productivity. All four types of compounds are endocentric, i.e. the verb acts as the head formally and semantically. As for the grammatical relation between the two constituents, the SUB and ATR relations are present while the CRD one is absent. We do not treat as compounds verbal VN formations like *portar noves* (bring news) 'to break the news to somebody', which we understand as lexicalized syntactic phrases. This subsection finalizes with Table 2.5 summarizing the results.

Some authors have denied the existence of verbal compounds in some Romance languages. For example, Rainer & Varela (1992) claim that there are no verbal compounds in Spanish, since they are not productive synchronically. Although verbal compounds in Catalan are not very productive, we will maintain that they do exist.

[NV]ᵥ compounds

Although the [NV]ᵥ compound has been argued to be unique to Catalan (e.g. Gavarró 1990b), it has been attested in other Romance languages like Spanish, e.g. *man+i+atar* (hand+and+tie) (cf. Rainer & Varela 1992), and Aude, e.g. *gorjo+badà* (throat+open) (cf. Klingebiel 1988).

This verbal compound has often been an object of discussion in the literature (e.g. Adelman 2002, Brunelli 2003, Gràcia & Fullana (G&F) 1999, 2000, Padrosa-Trias 2007a). The noun indicates inalienable possession and is usually a body part like *cor* 'heart' and *cama* 'leg', and the few cases in which the noun does not apparently indicate inalienable possession but inanimate or abstract entities (e.g. *aigua* 'water' and *terra* 'earth') are argued by Gavarró (1990b: 78-81) to be subject to 'The Non-Distinctness Constraint', according to which the noun outside the compound which the compound modifies and the noun inside the compound must have identical referents. Although the noun is generally interpreted as an argument of the verbal head (cf. e.g. Cabré & Rigau 1986, Duarte & Alsina 1986, Gavarró 1990b, Mascaró 1986, Pérez Saldanya et al. 2004, Wheeler 1977) - an affected theme in Gavarró's terms (cf. Rizzi's 1986 'theme 1') - we take a different position. Given that the compound can take an internal argument outside the compound, as in *El caçador cama+trencà l'ocell* (The hunter leg+broke the bird), we treat the noun inside the compound as a kind of modifier, which is in agreement with G&F's (1999: 246, 2000: 79) proposal, according to which the IPN is a modifier of the

complex predicate formed by the verb together with the possessor NP external to the complex verb. In the previous example, *cama* (leg) would modify *trencà l'ocell* (broke the bird) and a possible paraphrase could be 'to break the bird by the leg(s)'. G&F (2000: 245) compare the role of the IPN as a modifier of the verb to the 'Greek accusative'.[172]

The compound type under analysis gives further evidence for the model of grammar in which morphology and syntax are two separate components, with the compound being a morphological product. If the compound were the result of syntax, one would expect it to be left-headed given that Catalan syntax is left-headed. By contrast, morphology has both left-headed and right-headed structures, the type of compounds under study belonging to the latter group. It does not seem that the internal word order can be explained by appealing to a Latin source either. Oniga (1992: 101-103) provides only one example with non-transparent semantics: *tergi+versari* (back+turn) 'to hesitate'. On the basis of this example, not much can be concluded (e.g. whether the compound type in Latin was subjected to the same restrictions as the Catalan NV compound), but at least it suggests that this compound type was not that common, and that although some of the compounds in (75) could be traced back to Latin, most of them cannot. Note that Scalise (1992: 177) provides the Italian words *mano+mettere* (hand+put) 'to tamper with' and *croce+figgere* (cross+fix) 'crucify', which are Latin-based but regards them as unproductive.

All [NV]$_V$ compounds are endocentric SUB. The grammatical relation between the two components of the compound has already been discussed above. As for the endocentricity, the compound is a verb like the second element and tense marking is placed on it (formal head). Also, the compound denotes a type of action expressed by the verb (semantic head). Some examples follow:

(75)   aigua+batre (water+beat) 'to splash water', aigua+barrejar-se (water+mix+CL) 'to have waters of two rivers mix', ala+trencar (wing+break) 'to break the wing(s) (of an animal)', **cama+trencar** (leg+break) 'to break the leg(s)',

---

[172] Interestingly, real cases of NV sequences in which the N is interpreted as the internal argument of the V are found in child Catalan. Llinàs-Grau (1997) and Llinàs-Grau & Coll-Alfonso (2001) initially, and Tubau (2004) later, replicating the previous studies, show that children produce OV sequences at around the age of 2, although for a very short period of time (2 months approximately). The authors attribute the OV order, which is impossible in adult Catalan unless the object is focused or left-dislocated, to the fact that the child has not yet mastered the verbal morphology of Catalan and go through a bilingual stage. Note, though, that these early sequences are taken to be syntactic for the above-mentioned authors, not morphological (see the original works for details).

cap+alçar (head+lift) 'to lift (an object) up by the head', cap+ficar (head+put) 'to worry', cap+girar (head+turn) 'to turn upside down' and 'to change one's opinion', cap+trencar (head+break) 'to break something/somebody's head', cara+girar (face+turn) 'to turn the face' and 'to change opinion', coll+portar (neck+carry) 'to carry on one's shoulders', coll+tòrcer (neck+twist) 'to twist somebody/something's neck', coll+trencar (neck+break) 'to break somebody's neck' and 'to put forth a great effort', coll+vinclar (neck+bend) 'to bend somebody's neck', cor+bategar (heart+beat) 'to have the heart beat', cor+ferir (heart+hurt) 'to break somebody's heart', cor+glaçar-se (heart+freeze+CL) 'to get frightened', cor+nuar (heart+knot) 'distress', cor+prendre (heart+take) 'to captivate', cor+secar (heart+dry) 'to wither', pell+foradar (skin+pierce) 'to make a hole in something/somebody's skin', pell+obrir-se (skin+open+CL) 'to chap', pell+trencar-se (skin+break+CL) 'to have the skin or leather break', peu+calcigar (foot+step.on) 'to step on somebody's foot or to step on with one's foot', sang+cremar (blood+burn+CL) 'to get impatient', sang+glaçar (blood+freeze) 'to paralyse', tall+girar (blade+turn) 'to bend the blade (of a knife), terra+trémer (earth+shake) 'to have the earth quake', ull+ferir (eye+hurt) 'to hurt somebody's eyes (because of ugliness)', and ull+prendre (eye+take) 'to catch somebody's eye'.

Despite the apparent large number of compounds present in the language (75), Padrosa-Trias (2007a) concludes that [NV]$_V$ compounding is not very productive, contra Gavarró (1990b), G&F (1999, 2000), and Gràcia (2002). G&F (2000: 244), for instance, argue that Catalan speakers are able to understand novel instances of this compound. Such a statement needs to be taken with caution. By checking lists of N-V compounds with native speakers, Padrosa-Trias (2007a: 104) observes that "transparent N-V compounds can be understood but are very rarely produced (e.g. *cor+bategar* heart+beat) and that more opaque N-V compounds are often correctly understood but are rarely produced (e.g. *cor+ferir* heart+hurt 'to break somebody's heart')". Such results are in agreement with the findings in Adelman (2002), according to which compounds are more likely to be used in written form than in everyday conversation and by older speakers rather than by younger ones. Revealingly, nine subjects aged 15-19 are only 3% likely to use transparent N-V compounds in conversation. Note that these results are about the likelihood of using such compounds, and not about their real

use in conversation, which could be even lower. This is why Padrosa-Trias (2007a: 104-105) concludes that Catalan N-V compounds are "possibly a potentially unstable construction in the language" and that "if these forms were once productive, now they have fallen out of use, and possibly the ones which are still currently being used by some people are mostly being reanalysed as lexical items with only one root"; this explains why a few compounds have developed non-compositional semantics, like *cap+girar* (head+turn) 'to change one's opinion' and *sang+glaçar* (blood+freeze) 'to paralyse'. Padrosa-Trias questions the existence of some N-V compounds, for which she speculates that they may have been created (and stored in dictionaries) on the basis of their participial form, which is more common (e.g. *cor+secat* (heart+dry-PPLE) 'an embittered person' vs. $^?$*cor+secar* (heart+dry-INF) 'to wither') or is the only existing form in the language nowadays (*esma+perdut* (mood+lose-PPLE) 'disconcerted' vs. $^{*/??}$*esma+perdre* (mood+lose-INF) 'to lose heart'). Furthermore, novel instances of N+V compounds are more easily formed in their participial form than in their infinitival form: *ull+inflat* (eye+swell-PPLE) vs. $^?$*ull+inflar* (eye+swell-INF). (See Padrosa-Trias 2007a: 105-106 for these and other examples in Catalan; see also Rainer & Varela 1992 and Val Álvaro 1999, on the one hand, and Booij 2007, on the other, for the same findings in Spanish and Dutch respectively).

Notice that some [NV]$_V$ compounds are not subject to the restrictions mentioned above: the N does not indicate inalienable possession and, although in appearance it looks like an adverb, it acts as the internal argument of the verb. The verb determines the category of the compound and inflections are placed on it. Also, the action expressed by the compound is a type of action denoted by the verb. They are endocentric SUB compounds. There are only very few examples of this type and they are more commonly used in their participial form. Recall that forms with the same internal structure were treated in the subsection on nominal compounds, namely *ben+fer* (good+do) 'one's rectitude in their actions' and *ben+voler* (good+want) 'one's good will/love towards others' (cf. 72), but the result was nominal rather than verbal, due to their lexicalization.

(76)   **mal+dir** (bad+say) 'to say bad things (about somebody)'.

[AV]_V formations

There is no agreement with respect to the nature of the initial element of this compound. Some classifications do not make a distinction between adjectival and adverbial initial elements attaching to a verb, and [AdvV]_V and [AV]_V compounds are both treated as [AdvV]_V compounds (e.g. Adelman 2002, Cabré 1994, Mascaró 1986 and Padrosa-Trias 2007b). The fact that adverbs typically modify verbs and adjectives nouns suggests that when an item is an adjective in shape but is attached to a verb, it acts as if it were an adverb. On the other hand, other compounding classifications make a distinction between [AdvV]_V and [AV]_V compounds (e.g. Cabré & Rigau 1986, and G&F 2000) and place under the latter heading the compounds below (77) and sometimes some of the compounds we list as [AdvV]_V compounds in (78), like *car+comprar* (expensively+buy) 'to buy at an expensive price', and *car+vendre* (expensively+sell) 'to sell at an expensive price'. The double treatment of *car+comprar* and *car+vendre* may be due to the existence of two lexical items of identical form but of different category, one as an adverb and one as an adjective.

(77) **prim+filar** (thin+spin) 'to split hairs', prim+mirar (thin+see) 'to be really meticulous', and vil+tenir (vile+have) 'to vilify/underestimate'.

Notice that neither DIEC nor GDLC include *prim+mirar* (thin+see) 'to be really meticulous' but only its participial form *prim+mirat*, and although *prim+filar* (thin+spin) is included in both dictionaries, it is more commonly used as a participial form. *Vil+tenir* (vile+have) is not much used either as a conjugated verb or as a participial form. This leaves the group of [AV]_V compounds with (almost) no members, which we take as signalling the non-productivity (and probably the non-existence) of such a type. This is why [AV]_V compounds are included within parentheses in Table 2.5. They have been placed under SUB compounds since this seems to be the relation underlying the forms in (77). Contrast our conclusion with Gavarró's (1990b: 165). According to her, "the scarcity of new [AV]_V compounds" is due to factors of use rather than of grammaticality. She finds grammatical compounds like *ràpid+cantar* (quick+sing) 'to sing quickly' and *suau+tocar* (soft+touch) 'to touch softly', which we find ungrammatical. Recall that under the heading of adjectives, Gavarró includes both adjectives and adverbs.

AV compounds in other Romance languages like Italian have been claimed to be non-existent. Scalise (1992: 177) notes that forms like *gentile+parla* (kind+talk) and *caro+paga* (expensive+pay) are ungrammatical.

[AdvV]ᵥ compounds

Compound verbs formed by an adverb and a verb are not very productive, although there are a few examples of this type attested (78). Cabré & Rigau (1986) note that the adverbs tend to have a negative connotation: compounds with *mal-* 'badly' are common while those with *ben-* are rare, both of which being more easily found when adjoined to a participial verb than to a conjugated verb, e.g. *ben+vingut* (well+come-PPLE) 'welcome' vs. **ben+venir* (well+come-INF) (cf. subsection 2.3.2.3: adjectival [AdvA]ₐ compounds in Catalan).

Buenafuentes (2001-2002) identifies four different meanings of *mal* when it attaches to a verb in Spanish, which have exact counterparts in Catalan: quantitative (e.g. *mal+menjat*[173] (badly+eat-PPLE) 'to be undernourished'), privative (e.g. *mal+fiar-se* (badly+trust+CL) 'to mistrust'), intensive (*mal+ferir* (badly+wound) 'to wound badly') and qualitative (*mal+gastar* (badly+spend) 'to waste money'), of which we only take the latter to be a case of compounding (see Varela 2005: 79 who also treats the Spanish counterpart as a compound). The syntactic counterpart of *mal* in syntax is *malament*, which can only be interpreted with a qualitative meaning, i.e. 'in a bad manner'.[174]

(78)   car+comprar (expensively+buy) 'to buy at an expensive price', car+vendre
        (expensively+sell) 'to sell at an expensive price', mal+aconsellar (badly+advise)
        'to give bad advice', mal+acostumar (badly + get.used.to) 'to spoil (somebody),
        to get somebody into a bad habit', mal+baratar (badly+exchange) 'to squander',
        mal+casar (badly+marry) 'to marry (somebody) badly', mal+criar
        (badly+bring.up) 'to spoil (somebody)', mal+encaminar (badly+direct) 'to
        misdirect', mal+entendre (badly+understand) 'to misunderstand', **mal+gastar**
        (badly+spend) 'to waste money', mal+parlar (badly+speak) 'to speak ill of',

---

[173] Although the verb is in participial form, it just serves the purpose of signalling the quantitative reading that *mal* gives to the complex form.

[174] The Catalan examples are taken from Padrosa-Trias (2007b), who observes that despite *mal* having a qualitative meaning both in syntax and in the compound under examination, the syntactic counterpart of the compound generally has a wider range of interpretations.

mal+pensar (badly+think) 'to think badly', mal+tractar (badly+treat) 'to ill-treat', mal+vendre (badly+sell) 'to sell (something) cheap', menys+prear (less+praise) 'to underestimate', and menys+tenir (less+have) 'to undersestimate'.

[AdvV]$_V$ compounds are endocentric ATR compounds. The endocentricity comes from the fact that the verb is both the formal head (inflection is placed on the verb, which is also the element determining the category of the compound) and semantic head (the verb is a hyperonym of the compound). When it comes to classifying the compound with respect to the grammatical relation between the two elements, the task is always more difficult when the head is a verb than when it is a noun. In the case at hand, for instance, we interpret *mal+parlar* (badly+speak) as a person who speaks badly of somebody/something. Both an ATR relation and a SUB relation seem possible. When we look at the relation more closely, nonetheless, we can have the interpretation that the speaking (the speech content) is bad (ATR) but not necessarily that the (act of) speaking is carried out in a bad manner (SUB), hence the choice of ATR compounds (but see section 2.4 where the ATR-SUB distinction is further discussed).

[PV]$_V$ compounds

As was already discussed for nominal PN compounds, some authors (e.g. Mascaró 1986) treat such forms as compounds whereas others regard them as cases of prefixation (e.g. Cabré 1994, Cabré & Rigau 1986, Gavarró 1990b for Catalan; Lang 1992 for Spanish). For example, Gavarró (1990b: 166-167) argues that although the first element of a would-be compound looks like a preposition (e.g. *entre* 'between', *contra* 'against'), their pronunciation is not always the same as that of the independent counterpart: *contra+dir* [kɣntrəði] 'to contradict' vs. *contra tu* [kontrətu] 'against you'. The reduced vowels of complex words like *contra+dir* are taken by Gavarró as evidence that we are dealing with derivation. As for the examples with unreduced vowels, she argues that there are also prefixes with strong vowels. Concerning the semantics of the complex forms with *entre-*, for example, Gavarró identifies a meaning of reciprocity (e.g. *entre+mirar-se* (between+look+CL) 'to look at each other') and a meaning which is that of the base verb modified with the qualification 'slightly' (e.g. *entre+cavar* (between+dig) 'to dig superficially'). She contrasts the meaning of *entre-* with that of *entre* (the free counterpart) which can have a participative reading, which is

close to the reciprocal one but not quite the same. From such semantic differentiation, Gavarró concludes that forms with *entre-* must be derived words rather than compounds.

Some comments are in order here. The phonological and semantic arguments Gavarró provides only partially support a derivational analysis. We agree that reduced vowels may indicate derivation (or even compound lexicalization), but we think that although prefixes may have unreduced vowels, as Gavarró suggests, we view as equally valid the proposal that the first elements of compounds may also have unreduced vowels (cf. subsection 2.3.2). The semantic differentiation that Gavarró indicates between the free form *entre* and the bound form *entre-* suggests that the bound form can only be a prefix. However, Gavarró overlooks some crucial data: there are some cases in which the bound form has the same semantics as its independent counterpart. For example, the locative meaning of the preposition *entre* 'between' is identical to that found in *entre-* in *entre+posar* (between+put) 'to interpose'. In other words, it seems that two different cases of *entre*-forms should be distinguished: those which are unrelated to the free form (at least from a synchronic point of view) and should be considered part of derivation and those which are related to the free form and should be considered compounds.

In the group of PV compounds then, we will exclude those forms with a P expressing degree (which cannot be associated with the semantics of the corresponding free P). They will be taken as instances of derivation, namely prefixation, and not as an indication of compounding. Some examples which illustrate the point are as follow: *entre+cavar* (between+dig) 'to dig superficially', *entre+obrir* (between+open) 'to open halfway', *entre+tancar* (between+close) 'to close halfway', *entre+veure* (between+see) 'to see faintly, glimpse', *sobre+alimentar* (over+feed) 'to overfeed', *sobre+menjar* (over+eat) 'to overeat', *sota+excitar* (under+excite) 'to underexcite', and *ultra+congelar* (ultra+freeze) 'to deep-freeze'. In these cases, the prefix intensifies the action or state expressed by the verb.

According to Di Sciullo & Williams (1987) (cf. footnote 149, in subsection 2.3.1.2: verbal PV forms in English) the group of Ps which, we believe, form compounds would be further reduced. They make a distinction between function composition, which identifies affixation, and argument satisfaction, which may take place in compounding. If a P and a base together determine the final argument structure of the complex form, it means that it is a case of function composition and that we are

dealing with prefixation. This is rare among the prefixes that function like prepositions. We are aware of the case of *sobre-* only (79). Other prefixes which function like this are bound forms which do not have a free counterpart, like *des-* in *mentir/desmentir* (to lie/to deny a lie) and *re-* in *córrer/recórrer* (to run/to travel across), which we already considered to only be prefixes (80).

(79)   a. L'ocell vola (*el roure). 'The bird flies (*the oak tree)'
        b. L'ocell sobrevola *(el roure). 'The bird flies over *(the oak tree)'

(80)   a. En Joan sempre menteix (*coses). 'John always lies (*things)'
        b. En Joan desmenteix *(el que la Maria va dir) 'John denies *(what Mary said)'

Once the observations above have been taken into account, the conclusion is that there seem to be more restrictions when the P joins a V than when it joins a N to form compounds. The consequence is that there are only few PV compounds: their productivity is restricted. The Ps involved in the formation of PV compounds can indicate, among other meanings, location, as in *avant+posar* (before+put) 'to put something before', and reciprocity or relation between two entities, as in *entre+xocar* (between+crash) 'to crash into one another' and *contra+posar* (counter+put) 'to compare, to set against each other'. Some examples are given in (81). Notice that the compounds that take the reflexive clitic SE and indicate reciprocity or relation between two entites are more common without the P. The clitic seems to have taken over the function of the P.

(81)   avant+posar (before+put) 'to put something before', **contra+atacar** (against+attack) 'to counterattack', contra+batre (against+fight) 'to fight against somebody's fighting', contra+dir (counter+say) 'to contradict', contra+posar (counter+put) 'to compare, to set against each other', entre+besar-se (between+kiss+CL) 'to kiss each other', entre+creuar-se (between+cross+CL) 'to intersect, cross', entre+lligar (between+tie.up) 'to interweave', entre+matar-se (between+kill+CL) 'to kill each other', entre+mirar-se (between+look+CL) 'to look at each other', entre+xocar (between+crash) 'to crash into one another', sobre+sembrar (over+sow) 'to sow over a sown field', sobre+solar (above+sole)

'to put a new sole above the old one', sota+posar (below+put) 'to subordinate (somebody) to', and sots+arrendar (sub+rent) 'to sublet'.

Both the formal and semantic head coincide for all compounds, which explains the endocentricity of such compounds: they are verbs like the second element and the action expressed by the compound is a type of action denoted by the verb. The compounds seem SUB because the P is in a subordinating relation to the verb, although in some cases one could argue that they are ATR because the P seems to act as an adjective, especially in those cases where the P gives a reciprocal reading to the verb. As will be seen in the discussion section, the ATR-SUB division is illusory, since the two macro-types in B&S's (2005) classification can be subsumed under a unique macro-type. For the moment, all compounds in (81) are treated as endocentric SUB.

Table 2.5 summarizes the results of the present subsection. As usual, the table only includes those forms which we have considered as existing compounds in the language, and parentheses indicate that the compound type under consideration is not productive.

Table 2.5: Verbal Compounds in Catalan

	VERBAL COMPOUNDS					
	SUBORDINATE		ATTRIBUTIVE		COORDINATE	
	endocentric	exocentric	endocentric	exocentric	endocentric	exocentric
[NV]ᵥ	(cama+trencar) (75) (leg+break) (mal+dir) (76) (bad+say)					
[AV]ᵥ	(prim+filar) (77) (thin+spin) 'to split hairs'					
[AdvV]ᵥ			(mal+gastar) (78) (badly+spend)			
[PV]ᵥ	contra+atacar (81) (counter+attack)					

2.3.2.3 Adjectival compounds

Four types of adjectival compounds have been identified, all of which have an adjective (be it derived or underived) as the second constituent. The four compound types are as follows: NA, AA, AdvA and PA compounds. The three grammatical relations (SUB, ATR and CRD) are all present, although it does not mean that they are all instances of compounds. The forms with a CRD relation will be argued not to be compounds but phrases. The present subsection concludes with Table 2.6 gathering the results.

[NA]$_A$ compounds

This group of compounds includes those that have as a second element an adjective (e.g. *cama+curt* (leg+short) 'short-legged'), a participle (e.g. *llamp+ferit* (lightning+strike-PPLE) 'struck by lightning') and a deverbal adjective (e.g. *boca+badant* (mouth+open-PRES.PPLE) 'with one's mouth opened') since they all function like a simple adjective, but note that their productivity is not the same in the three cases. The compounds which are formed by an IPN and an adjective or a participle are the most productive ones: *cama+curt* (leg+short) 'short-legged' and *cara-xuclat* (face+suck-PPL) 'thin-faced' (cf. Mascaró 1986: 65). Those compounds with a deverbal adjective are the least productive ones: *boca+badant* (mouth+open-PRES.PPLE) 'with one's mouth opened'. Most of the compounds refer to animate entities (usually people), with the noun indicating a body part (e.g. *cara+xuclat* (face+suck-PPL) 'thin-faced'), but they can also refer to inanimate entities, although then they tend to be used as nouns only (e.g. *aigua+moll* (water+wet) 'marsh' (Mascaró 1986)).

The majority are endocentric SUB compounds (82a, b). The adjective functions as the formal and semantic head: it receives the plural and gender marking (the gender is not determined by the noun in the nonhead position but by the noun outside the compound that the complex adjective qualifies).[175] Regarding the semantics, the compound is understood as the property denoted by the adjective as it is applied to the noun in first position, i.e. the noun restricts the scope of attribution of the adjective (compare Gràcia 2002 for Catalan, Val Álvaro 1999 for Spanish). As for being SUB compounds, some nonhead nouns can be interpreted as an argument of the adjective, like *drogo+addicte* (drug+addict) 'drug addict', or as an argument of the deverbal

---

[175] Note that in Balearic Catalan the adjective agrees in gender and number with the noun inside the compound: *un noi llengua+llarga* ([a boy$_{masc}$ [tongue$_{fem}$+long$_{fem}$]]) in Balearic Catalan vs. *un noi llengua+llarg* ([a boy$_{masc}$ [tongue$_{fem}$+long$_{masc}$]]) in continental Catalan 'a foul-mouthed boy' (cf. Moll 1975: 246).

adjective like *boca+badant* (mouth+open-PRES.PPLE) 'with one's mouth opened'. If the participial suffix were separated from the verb and were treated as the head of the word, the compound would still be SUB: 'one who opens their mouth'. Other nouns are interpreted as a modifier of the head adjective: *cama+curt* (leg+short) 'short-legged' is interpreted as *curt de cames* (short of legs)[176] or *llamp+ferit* (lightning+strike-PPL) 'struck by lightning' is understood as *ferit per un llamp* 'struck by thunder'. We were able to find one example which is not SUB, but ATR: *pal+plantat* (stick+plant-PPL) 'still as a statue' (82c).

The SUB compounds below have been divided into two groups. The first one (82a) includes compounds with an IPN, the possessor of which is the noun outside the compound which is qualified by the compound adjective: *un noi cama+curt* (a boy short+leg) 'a short-legged boy' (cf. verbal VN compounds like *cama+trencar* (75), which also include IPNs). The first noun is always singular, even though it may denote some plurality: *peu+gròs* (foot+big) 'big-footed' and *cella+junt* (eyebrow+joint) 'having joint eyebrows' (for more discussion on the issue of inalienability concerning this compound type, see e.g. Gràcia 2002 for Catalan and Sánchez López 2003 for Spanish). The second group (82b), by contrast, includes compounds with no IPN, although they are also attributed to an entity outside the compound. Neither group is used for the creation of new vocabulary in specialised vocabulary.

(82) a. ala+caigut (wing+fall-PPL) 'with the wings fallen' and 'feeling down', ala+llarg (wing+long) 'long-winged', ala+ferit (wing+hurt-PPL) 'hurt-winged', anca-rossegant (haunch+drag-PRES.PPLE) 'downhearted', barba+blanc (beard+white) 'white-bearded', boca+badant (mouth+open-PRES.PPLE) 'with one's mouth opened', boca+moll (mouth+wet) 'indiscreet', boca+tort (mouth+crook-PPLE) 'crooked-mouthed', **cama+curt** (leg+short) 'short-legged', cama+llarg (leg+long) 'long-legged' and 'wading bird', cama+lluent (leg+shine-PRES.PPLE) 'with shiny legs', cap+baix (head+low) 'to be sad', cap+gròs (head+big) 'big-headed' and 'tadpole', cara+ample (face+wide) 'wide-faced', cara+prim (face+thin) 'thin-faced', cara+rodó (face+round) 'round-faced', cara+xuclat (face+suck-PPL) 'thin-faced', cella+junt (eyebrow+joint) 'having joint eyebrows', coll+ample (neck+wide) 'wide-necked', cua+curt (tail+short)

---

[176] See Sánchez López (2003) for subtle semantic differences with the paraphrases which are often attributed to the adjectival NA compound (i.e. 'A of N', and 'of NA').

'short-tailed', cua+llarg (tail+long) 'long-tailed', cul+gròs (bottom+big) 'big-bottomed' and 'a mushroom: *Amanita ovoidea*', front+ample (forehead+wide) 'wide-foreheaded', galta+plè (cheek+plump) 'plump-cheeked', llavi+gròs (lip+thick) 'thick-lipped', llengua+llarg (tongue+long) 'foul-mouthed', mà+llarg (hand+long) 'long-handed', panxa+content (belly+happy) 'laid back', pell+roja (skin+red) 'a red skin, an American Indian', peu+gròs (foot+big) 'big-footed', un pit+roig (chest+red) 'a robin', and ull+blau (eye+blue) 'blue-eyed'.

b. clau+passat (nail+pass-PPL) 'weak due to an illness', creu+clavat (cross+fix-PPL) 'crucified', drogo+addicte (drug+addict) 'drug addict', fe+faent (faith+make-PRES.PPL) 'worthy of faith', gel+cuit (ice+cook-PPL) 'iced', llamp+ferit (lightning+strike-PPL) 'struck by lightning', mal+dient (bad+say-PRES.PPL) 'relating to somebody who says bad things (about somebody)', sol+cuit (sun+cook-PPL) 'sunburnt', and tot+poderós (all+powerful) 'the Almighty'.

c. **pal+plantat** (stick+plant-PPLE) 'still as a statue'.

Some NA compounds may undergo conversion and be treated as nouns. This is the case of most lexicalised compounds: *un cama+llarg* (a leg+long) 'wading bird', *un cap+gròs* (a head+big) 'tadpole', *un cul+gròs* (a bottom+big) 'a mushroom: *Amanita ovoidea*' (cf. Cabré & Rigau 1986: 143, Mascaró 1986: 73, Gavarró 1990b: 172) (cf. subsection 2.3.2.1: nominal NA compounds). Others are still adjectives but also adopt a metaphorical reading: *ala+caigut* (wing+fall-PPLE) 'feeling down', *anca-rossegant* (haunch+drag-PRES.PPLE) 'downhearted', *llengua+llarg* (tongue+long) 'foul-mouthed' and *panxa+content* (belly+happy) 'laid back'.

Also, note that the compounds whose second element is a participle or a deverbal adjective are more commonly used as participial forms than verbal forms and sometimes they are the only existing form. Compare *cara-xuclat* (face+suck-PPL) 'thin-faced' with [??]*cara-xuclar* (face+suck-INF) (cf. subsection 2.3.2.2: verbal NV compounds). Gràcia (2002: 813) attributes this fact to aspectual factors.

Adjectival NA compounds are not unique to Catalan. They are also present in other Romance languages, although in Spanish, for example, they seem to have a different structure: N+i+A (N+and+A), as in *oj(o)+i+negro* (eye+and+black) 'black-

eyed'.[177] If the intervening vowel is taken as a linking morpheme whose only function is to link the two compounding elements, the compound in Spanish parallels that in Catalan (cf. García Lozano 1978, Val Álvaro 1999, Sánchez López 2003): the linking element in Spanish does not change the category of either constituent of the compound and the compound could be regarded as the union of a noun and an adjective. Like in Catalan, the compound in Spanish would be endocentric SUB with the adjective being the head.

On the other hand, if the –i- vowel is seen as a derivational morpheme which changes the first noun into an adjective, then the compound is the sum of two adjectives, a possibility which is not available to Catalan (cf. Clements 1992, Gil Laforga 2006). For example, Gil Laforga (2006), following H&K (1993, 2002), defends such a position and argues that the denominal adjective is the head of the compound. On this analysis the second adjective modifies the underlying noun in the first position. Note, though, that Gil Laforga's (2006) analysis faces some problems: she adopts H&K's (2002) idea that conflation is merge, and conflation is understood as copying the phonological material of the sister head into the higher phonological empty head. The immediate consequence of conflation is that material which is not included in the lower sister head will not be copied to the higher head. According to Gil Laforga's (2006: 35-37) proposal, though, the specifier of a phrase moves into the head position of the higher phrase, an illicit movement if we want to observe H&K's (2002) idea that conflation equals merge.[178] (See Padrosa-Trias 2007a for a review of some problems associated with the conflation-merge equation when applied to some compounds).

---

[177] The vowel intervening between the two compounding elements was not present initially. It is hypothesized that the vowel was brought about by imitation of the Latin counterpart, as in *barb+i+rasus* (cf. García Lozano 1978, Sánchez López 2003). Sánchez López (2003: 164-166) further speculates that the role of the vowel nowadays is to mark the subordinating relation of the noun with respect to the adjective, which he takes as the head of the compound. See Gil Laforga (2003) for a summary of the different diachronic and synchronic views on the vowel –i-.

[178] The tree structure proposed by Gil Laforga (2006: 37) for *oj(o)+i+negro* (eye+and+black) 'black-eyed' is given in (i). PosP (Possessive Phrase) represents a possession relation and δP stands for a predication relation.

(i)

PossP
/\
Poss
/\
[ ]i   δP
/\
ojo δ
/\
δ   negro
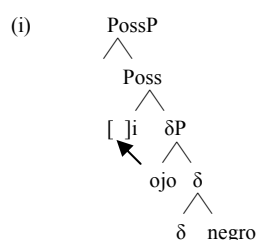
Not everybody agrees on the endocentric nature of the compound in Spanish. In this respect, Rainer & Varela (1992: 133) observe that "while right-headed compounds are available for head-operational derivation (*droga+dicción* (drug+addiction)), *clar+i+videncia* (clair+voyance), etc.)", N+i+A compounds cannot undergo derivational processes of this kind (e.g. **lengü+i+largura* (tongue+and+length), **cuell+i+cortedad* (neck+and+shortness), which they take as meaning that the compound is exocentric. Their claim is that *droga+dicción* (drug+addiction) is derived from *droga+dicto* (drug+addict) and *clar+i+videncia* (clair+voyance) from *clar+i+vidente* (clair+voyant), which is possible because the second constituent is the head. The impossibility for N+i+A compounds (e.g. *lengü+i+largo* (tongue+and+long), *cuell+i+corto* (neck+and+short)), to undergo such derivational process would remain a mystery in an endocentric approach if such an operation were a real one. The fact that compounds like *droga+dicto* (drug+addict) and *clar+i+vidente* (clair+voyant) can undergo suffixation does not necessarily mean that the head is on the right, though. These compounds are not native compounds: *drogadicto* is taken from English and *clarividente* from Latin, and are most probably perceived as simplex words which can undergo suffixation as simplex words do.

For the restrictions of this compound type in Spanish, see García Lozano (1978) and Val Álvaro (1999), and for the restrictions of this compound in Catalan, see Mascaró (1986: 65-66) and Pérez Saldanya et al. (2004: 265), among others.

[AA]$_A$ compounds

Forms of this type can initially be divided into four subgroups. According to B&S's (2005) compounding scheme, the forms in (83a) and (83b) would both be CRD compounds, and they would differ in being endocentric and exocentric respectively. Regarding the compounds in (83c) and (83d), they would be exocentric ATR compounds. After presenting how the four subgroups would be analysed in B&S's (2005) compounding classification, we will show that there is no difference between the forms in (83a) and (83b): the coordinate relation will be argued to be syntactic, which can be part of a compound when inserted in its non-head position. As for the compounds in (83c) and (83d), we will show that they are endocentric SUB compounds.

(83)   a. agre+dolç (sour+sweet), anglo+català (English+Catalan), físico+químic (physical+chemical), greco+llatí (Greek+Latin), greco+romà (Greek+Roman),

hispano+argentí (Hispano+Argentinian), sord+mut (deaf+mute), and teòrico-pràctic (theoretical+practical).

b. (diccionari) anglès-català ((dictionary) English+Catalan), **(relacions) catalano-occitanes** ((relations) Catalan+Occitan), (diccionari) francès-espanyol ((dictionary) French+Spanish), (prefix) greco-llatí ((prefix) Greek+Latin), and (tractat) hispano-americà ((treaty) Hispano+American)).

c. ben+aventurat (well+ventured) 'blessed', ben+cossat (good+bodied) 'having a well-proportioned body', ben+humorat (good+humoured) 'good-humoured', mal+carat (bad+faced) 'surly', mal+dentat (bad+toothed) 'having uneven and not orderly arranged teeth', **mal+humorat** (bad+humoured) 'bad-tempered', and mal+intencionat (bad+intentioned) 'ill-intentioned'.

d. alt+i+sonant (high+and+sound-PRES.PPLE) 'grandiloquent, high-sounding', clar+i+vident (clear+and+see-PRES.PPLE) 'clear-sighted, clairvoyant', nou+nat (new+born-PPLE) 'newborn', prim+mirat (thin+look-PPLE) 'of somebody who is really meticulous', and **ver+semblant** (true+seem-PRES.PPLE) 'credible, plausible'.

Concerning the adjectives of the compounds in (83a), a distinction should be made between those compounds whose first element ends in –o, which are very productive, and those that do not finish in –o, which are not productive (Gràcia 2002: 817). Despite the fact that those adjectives whose first element ends in –o have the appearance of being learned compounds, we do not adopt this view, which explains their inclusion here. The compounds consist of two native adjectives and these are not technical terms. The only feature that these words take from learned compounds is the linking vowel –o-. Regarding the types of adjectives that can appear in the compound, Val Álvaro (1999: 4808) distinguishes three types for the Spanish counterparts of the forms in (83a): adjectives which refer to colours (e.g. *azul+violeta* (blue+violet)), to nationalities (e.g. *anglo+americano* (Anglo+American)) and to several lexico-semantic domains (e.g *político-social* (political+social)). Recall that we regarded the names of colours as NN, and not as AA, compounds.

There are several proposals as to where the (semantic) head is. Mascaró (1986: 73) argues that the compounds under analysis are endocentric, the two constituents being symmetrical; Gavarró (1990b: 173-174) treats compounds like *agre+dolç* (sour+sweet) as right-headed; and Cabré & Rigau (1986: 144-145) maintain that they are non-headed. To our understanding, the two adjectives are both semantic heads. For example, if we have *una salsa agre+dolça* (a sauce sour+sweet), the sauce is both sour and sweet on equal terms. Formally, for those compounds which are not invariable, number and gender marking is superficially placed on the second constituent, which we interpret as having scope over the two elements if the semantic double-headed analysis is right. Val Álvaro (1999: 4771) reaches the same conclusion for Spanish: the apparent inflection on the second constituent has scope over the two elements in words like *sordo+mudo+s* (deaf+mute-PL). The endocentricity of these compounds and the CRD relation between the constituents are thus derived. (Recall that Lieber 2008 distinguishes between 'simultaneous' and 'mixture' compounds).

Although the compounding elements could in principle be interchangeable due to the CRD relation between them, they are not. Rainer & Varela (1992: 131) identify some restrictions, or "preference rules" as they call them, for the Spanish counterparts of the compounds in (83a): (i) bound constituents do not appear in final position, (ii) the longer constituents are usually placed in second position, and (iii) if one element ends in −*o* and the other in a consonant, the first one tends to occur in first position.

As for the compounds in (83b), following B&S's (2005) scheme, they are exocentric CRD compounds (in Lieber's 2008 terms: they can be 'relationship' and 'collective'). The two elements of the compounds are attributed to an external entity. Whether formal marking is placed on the two constituents or on just one is irrelevant for the issue of exocentricity, because the semantic head is located outside the compound. The CRD relation is mostly understood as having the conjunction 'and' (e.g. relations between Catalan *and* Occitan people in *relacions catalano-occitanes* (relations Catalan+Occitan)), or a directional element (e.g. a direction *from* English *to* Catalan in *diccionari anglès-català* (dictionary English+Catalan)).

Having said that, we want to defend the view according to which there is no difference between the coordinated elements in (83a) and (83b). We believe that they have both been created by syntax and hence are not compounds, since for us a true

coordinate relation can only be established in syntax. A coordinate structure can, nonetheless, become part of a compound when it is inserted in the non-head position of a compound, which is the case of the forms in (83b).

If the forms in (83a) and (83b) are really the same, the traditional association of the forms in (83a) with endocentricity and the forms in (83b) with exocentricity must be explained in different terms. We contend that the endocentricity/exocentricity distinction is illusory for the forms at issue, because, by its very nature, an adjective denotes a property which must be attributed to an entity and that is the case for the forms in both (83a) and (83b). This explains why any of the forms in (83a) needs a noun outside the coordinate structure: *una **salsa** agre+dolça* (a **sauce** sour+sweet), like any of the forms in (83b). We further believe that it is the noun outside the coordinate structure that determines the resulting semantics of the complex word and not the coordinate structure itself, as has been suggested by Gràcia (2002: 817), who argues that the same compound (i.e. coordinate structure for us) can have two different interpretations, which can be distinguished by the presence or absence of a hyphen between the two compounding elements. In our view, such semantic differentiation should be attributed to the properties of the noun located outside the coordinate structure. For example, in virtue of having *diccionari* in *un diccionari anglès-català* (a dictionary English+Catalan), we understand that there is a direction from English to Catalan and not a blurred mixture of English and Catalan. In contrast, by having *noi* in *un noi anglo+català* (a boy English-Catalan), we understand that the person is half English and half Catalan, and that there is no direction implied. One might think that it is the different allomorphy of the word *English* in Catalan (*anglès* vs. *anglo*) which gives the different readings, but there are cases where the same exact form can also be involved in two different readings. Consider *anglo+americà* (Anglo+American): *relacions anglo+americanes* (relations English+American-PL) can be understood as relations which have been established between England and America but *una persona anglo+americana* (a person English+American) will be understood as an American person who is of English descent. Such difference in interpretation can only be attributed to the presence of *relacions* vs. *persona*.

In short, the forms in (83a) and (83b) behave the same semantically and formally. Given the coordinate relation between the two adjectives, formal marking is expected to be on both of them, but this is hardly ever the case. Let us consider why. Regarding the forms in the first group, plural and gender marking is placed at the end.

Formal marking cannot appear after the first adjective when it ends in –*o* because such an element, we believe, prevents other formal markers from appearing in this position. As for *agre+dolç* (sour+sweet) and *sord+mut* (deaf+mute), the combination of the two adjectives has become fixed, conventionalised and ultimately seen by speakers as a simplex adjective, which explains why gender and plural marking is placed at the end. Regarding the forms in the second group, similar results are found. The presence of the linking vowel –*o* at the end of the first adjective in some forms prevents it from having number and gender agreement, which as a result is only placed at the end of the coordinate structure. As for compounds like *diccionari francès-alemany* (dictionary French+German), the coordinate structure tends to be invariable when the head *diccionari* is pluralized, because what is at stake is the translation of one language into another, and not the translation of more than one language into others. However, some speakers also place plural inflection on both adjectives and some others only at the end of the coordinate structure. If instead of using an adjective ending in –*o* (in its learned version), we use it in its native form, and the head of the compound is pluralized, then the adjectival coordinate structure tends to be inflected for plurality at the end, although plural inflection on both adjectives is not totally excluded either: *unes llegendes catalana-angleses* (some legends Catalan-SG+English-PL) and *unes llegendes catalanes-angleses* (some legends Catalan-PL+English-PL) are both possible. The interpretation can be that the legends exist both in the Catalan and English traditions, or that there are words from the two languages. The mixed results of plural marking seem to suggest that agreement on adjectives is non-interpretable (Chomsky 1995a), but note that speakers are uneasy about the coordination of two adjectives, the first one of which does not end in –*o*. Speakers prefer to have the first adjective ending in –*o* or an overt conjunction between the two adjectives (e.g. *and*). Similar findings are found in Spanish (cf. Rainer & Varela 1992: 132, Val Álvaro 1999: 4810-4812). In short, an expression like *diccionari francès-alemany* (dictionary French+German) would be an endocentric ATR compound and a structure like *relacions anglo+americanes* (relations English+American) would be an endocentric SUB compound. Recall that they are compounds not by virtue of the coordinate structure (which we regard as syntactic) but by virtue of the relation established between the nominal head (e.g. *diccionari* and *relacions*) and the non-head, which happens to be a coordinate structure of two adjectives (e.g. *francès-alemany* and *anglo+americanes*). Consequently, this NA compound type is not included in Table 2.6, where adjectival compounds are listed, but

in Table 2.4, which contains nominal compounds. Only the SUB type is exemplified. (The ATR/SUB distinction is taken up in the next section).

The compounds in (83c) can be considered endocentric SUB, although an ATR relation is also present (i.e. the adjective in first position modifies the noun underlying the derived adjective in second position). Let us consider *mal+humor+at* (bad+humour+ed): although the relation between *mal* and *humor* is ATR (the humour is bad), a relation which could lead one to think that the compound is ATR, one has to bear in mind that such an ATR relation is subordinated to a head outside the AN structure, i.e. the ornative suffix, thus deriving the SUB relation of the compound as a whole. The ornative suffix attaches to the noun despite having scope over the whole compound (a bracketing paradox). Gender and plural marking is placed on the ornative suffix (hence, the endocentricity of the compound). Observe that most first elements are either *ben* or *mal* and that some compounds have become lexicalized: *ben+aventurat* (well+ventured) 'blessed'. The absence of other adjectives in first position leads us to question the nature of *ben* and *mal*, which could be acquiring the status of prefixes. For this reason, these compounds are put within parentheses in Table 2.6. These compounds have the same internal structure as the English compounds listed in (55d), for example.

Finally, the compounds in (83d) only exist as adjectival participles, that is, the verbal base from which they seem to derive does not exist (cf. [NV]$_V$ compounds in the subsection on verbal compounds). Most of the underlying verbs are verbs of perception and select the adjective in first position. This compound type is not productive and some forms have an –*i*-, like *clar+i+vident* (clear+and+see-PRES.PPLE) 'clear-sighted, clairvoyant' and *alt+i+sonant* (high+and+sound-PRES.PPLE) 'grandiloquent, high-sounding'. Such forms resemble the Latinate compounds which also have an –*i*-, which might indicate the dependence relation of the first element on the second one, e.g. *barb+i+rasus* (cf. footnote 177), and in fact these forms are Latin-based, with *clar+i+vident* being introduced into the language via French. Authors like Rainer & Varela (1992: 132) and Val Álvaro (1999: 4822) treat as adverbs the initial element of the same compounds in Spanish (e.g. *alt+i+sonante* (high+and+sound-PRES.PPLE)). The question of whether the first element of these two forms is adverbial or adjectival is of not much importance, since these forms are probably not decomposable by speakers and no new forms can be created following the same pattern. In addition, recall that

although we keep the labels 'adjective' and 'adverb' separate, we regard adverbs as derived from adjectives. As for the rest of the forms in (83d), although unproductive, the nature of the first element is adjectival, hence the inclusion of these forms here.

These compounds could be qualified as endocentric SUB. For those compounds which are still decomposable, the semantic head is the participial suffix: the adjective and the underlying verb are semantically subordinated to the suffix (SUB relation), which attaches to the verb but has scope over the adjective and the verb together. For example, *prim+mirat* ((thin+look-PPLE) 'of somebody who is really meticulous') takes formal marking (gender and number) at the end, which we interpret as being on the suffix. Thus, the endocentricity of this compound is explained. Some compounds are invariable (e.g *ver+semblant* (true+seem-PRES.PPLE) 'credible, plausible'). The relation between the adjective and the verb can be seen as ATR: the adjective assigns an attribute to the verb (Gràcia 2002: 814-815, Cabré & Rigau 1986: 144-145).

[AdvA]_A compounds

These adjectival compounds are deverbal formations whose allegedly verbal base is non-existent: compare *mal+endreçat* (bad+arrange-PPLE) 'arranged not in an orderly manner' with **mal+endreçar* (bad+arrange-INF). These complex forms have received varied treatments in the literature. For example, they are not included in Cabré & Rigau (1986) and in Gavarró (1990b) these complex forms are derived from phrases which have become lexicalized and now are idiosyncratic in meaning.

(84)   ben+estant (well+be-PRES.PPLE) 'well-to-do', **ben+parlat** (well+speak-PPLE) 'of somebody who speaks without swearing', ben+vingut (well+come-PPLE) 'welcome', mal+endreçat (bad+arrange-PPLE) 'arranged not in an orderly manner', mal+sonant (bad+sound-PRES.PPLE) 'rude (word)', prop+dit (near+say-PPLE) 'just said', prop+passat (near+pass-PPLE) 'recent', prop+vinent (near+come-PRES.PPLE) 'next'.

Although all forms are a sequence of an adverb and an adjective superficially, two types can be distinguished in terms of their internal structure. Both types of compounds can be considered endocentric SUB. In compounds like *ben+parlat* (well+speak-PPLE), the participial suffix can be taken as the semantic and formal head. The suffix attaches to the verb formally although it has scope over the adverb and verb together. The resulting

complex adjective agrees in number and gender with the noun outside the compound: *un noi ben+parlat* (a boy well+speak-PPLE) is *un noi que parla bé/sense paraulotes* (a boy who speaks well/without swearwords). There are other compounds in which the participial suffix has fused into the verb in such a way that it is no longer perceivable. In such cases, the participial adjective is treated as underived and is the formal and semantic head. Formal marking is placed on the participial adjective, which is a hyperonym of the compound (endocentricity): *un calaix mal+endreçat* (a drawer bad+arrange-PPLE) is *un calaix que està endreçat malament* (a drawer which is arrange-PPLE badly). As for the SUB relation, the adverb describes how the action in the participle is carried out.

This type is not very productive (cf. Gràcia 2002: 815) and some compounds are instances of lexicalization, as in *ben+estant* (well+being) 'well-to-do', and *ben+vingut* (well+come-pple) 'welcome'.

[PA]ₐ compounds

This compound type is not as common as nominal PN and verbal PV compounds. Some prefixes have developed a degree meaning which is not present in their free counterparts, and hence they are considered prefixed words and are removed from PA compounds (85). Some examples of prefixed words include *sobre+bò* (above+good) 'very good', *sobre+plè* (above+full) 'very full', and *ultra+lleuger* (ultra+light) 'very light', which denote a certain degree of the property expressed by the adjective, and not a location as the independent preposition indicates. Recall that authors like Cabré (1994) and Cabré & Rigau (1986) treat all PA forms as cases of prefixation. Lexicalized PA compounds have not been taken into account either (e.g. *avant+guardista* 'avant-garde').

Regarding the alleged PA compounds in (85), some have participles as second members, like *avant+dit*[179] (before+say-PPLE) 'previously mentioned', and *sota+escrit* (under+write-PPLE) 'undersigned', the latter deriving from the verb *sota+escriure* (under+write) 'to sign at the foot (of a document)'. They are dubious cases of PA compounds, because the first element can also be analysed as an adverb, which is the treatment adopted by Scalise (1992: 178, fn. 6) for Italian compounds like *sopra+citato* 'above mentioned'.

---

[179] *Avant+dit* (before+say-PPLE) is included in GDLC but not in DIEC.

In the rest of the PA compounds, the adjectival base has an underlying noun, which acts as the complement of the P: *ultra+marí* (ultra+marine) (<*mar* 'sea'), *contra+natural* (against+natural) (<*natura* 'nature'). It is a case of a bracketing paradox: semantically the P and the noun go together and the adjectival suffix takes scope over them, but formally the adjectival suffix attaches to the noun. The suffix is then the formal and semantic head: it determines the category of the compound and, as already noted, it takes scope over P+N. The relation between the P and the noun is one of subordination, and the outermost relation, i.e. the relation between the suffix and the P+N, which is the one determining the compound type, also seems to be subordination. For example, *una acció contra+natural* (against+natural) is understood as an action that goes against nature (SUB compound). As can be observed from the examples below, PA compounds have some of the meanings already discussed for PN and PV compounds: location (*ultra+marí* (ultra+marine) 'overseas') and opposition (*contra+natural* (against+natural) 'anti-natural').

Once the dubious cases (the forms with a past participle in second position) are removed from (85) and we take into account the fact that many of the other forms involve the preposition *contra*, we conclude that such a compound type is not very productive.

(85)   avant+dit   (before+say-PPLE)   'previously   mentioned',   **contra+natural** (against+natural)   'anti-natural',   contra+produent   (against+productive) 'counterproductive',   contra+reformista   (against+reformist)   'anti-reformist', sota+escrit (under+write-PPLE) 'undersigned', sota+signat (under+sign-PPLE) 'undersigned'and ultra+marí (ultra+marine) 'overseas'.

This brings to an end the subsection of adjectival compounds in Catalan, whose results are gathered in Table 2.6. As before, the table only contains compounds and the low productivity of some compound types is indicated by means of parentheses.

Table 2.6: Adjectival Compounds in Catalan

	ADJECTIVAL COMPOUNDS					
	SUBORDINATE		ATTRIBUTIVE		COORDINATE	
	endocentric	exocentric	endocentric	exocentric	endocentric	exocentric
[NA]_A	cama+curt (82a) (leg+short)		(pal+plantat) (82c) (stick+plant-PPLE)			
[AA]_A	(mal+humorat) (83c) (bad+humoured) (ver+semblant) (83d) (true+seem-PRES.PPLE)					
[AdvA]_A	(ben+parlat) (84) (well+speak-PPLE)					
[PA]_A	(contra+natural) (85) (against+natural)					

2.4 Discussion and conclusion

Chapter 2 (section 2.1) started with some discussion about the existence of heads in morphology, which was confirmed (contra Zwicky 1985, Bauer 1990, and Anderson 1992). Some arguments against the postulation of morphological heads derived from the attempt to apply the same set of criteria to identify the head in syntax to morphology. We saw that heads exist both in syntax and morphology, but that they are not identical. Consequently, one should not try to determine the syntactic and morphological heads by means of the same criteria. The notion of head has played a fundamental role in the classification of compounds in English and Catalan.

Section 2.2 was divided into two subsections: subsection 2.2.1 was devoted to identifying the nature of the compounding elements in English and Catalan. We concluded that English compounds can have (i) a root, a lexeme or a phrase in the non-head position, and (ii) a root or a lexeme in the head position, roots only being used in neo-classical compounds. Catalan compounds can have (i) a root, a stem, a lexeme or a phrase in the non-head position, and (ii) a root, a stem or a lexeme in the head position, roots also being used in the case of neo-classical compounds only.

Subsection 2.2.2 provided a brief overview of several types of compound classifications, which included a transformational account of compounds (Lees 1960), classifications based on a set of semantic classes (Downing 1977, Hatcher 1960, Levi 1978), classifications based on the syntactic categories of the input and output categories (Carstairs-McCarthy 2002, Plag 2003) and a mixture of the previous classifications (Adams 1973, Bauer 1983, 2003, Booij 2005). Classifying compounds proved to be quite a difficult task: none of the classificatory schemes was satisfactory enough. Finally, what looked to be the most promising classification of compounds currently available, namely B&S's (2005) classification, was explored. It is intended to be universal and based on consistent criteria. It provides three macro-types of compounds, which are based on the grammatical relation between the two constituents: SUB, ATR and CRD. They are in turn divided into two subtypes: endocentric vs. exocentric, a distinction which was initially based on a rough notion of head. Such a notion is revised in Scalise & Guevara (2006), which results in a better understanding of endocentricity vs. exocentricity. B&S acknowledge that the classification, as it stands

now, contains rough subdivisions. A way of refining their classification is by adding another layer of analysis to their two levels, to which section 2.3 was devoted.

Section 2.3 provides a thorough study of the compounds available in present-day English and Catalan and classifies them using the syntactic categories of the input and output categories. Such categorial-based classification is incorporated into B&S's classificatory scheme. The addition of this third level of analysis allows compounds to be further distinguished and makes cross-comparison of languages easier. The classifications of compounds in English and Catalan represent an improvement over the classifications available so far. Note that B&S exemplified their 2-level classification with compounds from English whose constituents were adjectives and nouns only. As we have seen, English compounding can make use of more input categories. As for Catalan, to our knowledge, there is no classification as sophisticated as the classification developed here. In addition, our third level of analysis will allow us to corroborate in the next chapter Snyder's (2001) hypothesis about the alleged correlation between resultatives and productive compounding in a language, which can be based on the category of the input elements (at least in one reading).

B&S's classificatory scheme was our starting point. After presenting the compounds in English and Catalan according to their classification, which incorporated the input/output categories as a third level of analysis, we departed from it in substantial ways during the course of subsections 2.3.1 and 2.3.2. We denied the existence of exocentric compounds and CRD compounds: all compounds are endocentric and what are generally called CRD compounds are cases of asyndetic syntactic coordination and not compounds. Coordinate structures, though, can be reinterpreted as compounds when one element is taken as the head and the other as the non-head (e.g. *player coach*). Coordinate structures can also become part of a compound when they are inserted in the non-head position of the compound (e.g. *the mind-body problem*). The denial of CRD compounds is a departure from most studies on compounding (e.g. B&S 2005, Pérez Saldanya et al. 2004, Olsen 2001, 2004, Val Álvaro 1999, among many others). Recall that, for some authors, CRD compounds only exist in some languages. For example, Olsen (2001) argues that 'copulative compounds' (using her terminology) exist in Germanic, but not in Romance or Sanskrit where they are syntactic configurations in her opinion. We share Olsen's view only partially: we believe that if there is a truly coordinate structure, it cannot be a compound by itself in any language.

As for SUB and ATR compounds, we noted the difficulty distinguishing between them on more than one occasion in this chapter. For example, recall that *player coach* (36a) can be treated both as a SUB compound and as an ATR compound by native speakers: 'a coach that plays with the team' (SUB) and 'a coach who is also a player on the team' (ATR). Similarly, compounds with the same structure (the non-head being a coordinate phrase) (cf. 55b) can be treated as SUB, like *a public-private partnership* 'a partnership *between* public and private organisations', and as ATR, like *a cruel-compassionate expression* 'an expression which *is* cruel but compassionate'. In a similar fashion, the Catalan compounds in (64a) can be treated as ATR/SUB compounds: e.g. *vagó restaurant* (wagon restaurant) 'dining car' can be seen as a wagon which *is* like a restaurant (ATR) and as a wagon {*with/which has or contains*} a restaurant (SUB). What we believe is at stake in all cases of compounding is the head vs. non-head relation.[180] Accordingly, the distinction between attributive compounds and subordinate compounds is irrelevant for this purpose, since they are both subsumed under the same relation, and whether the non-head is a kind of attribute or complement to the head follows from the (context and) semantics of the compounding elements. In short, all compounds are based on the same structure:

(86)   a.   b.

       non-head   head   head   non-head

Regarding productive compounds, English compounding is represented by means of (86a) while Catalan compounding is mainly characterized by (86b) although some

---

[180] Our conclusion (i.e. all compounds are based on a non-head vs. head pattern) seems to comport well with Di Sciullo's (2005, 2007) Asymmetry Theory. Di Sciullo (2005: 13) proposes *the Strict Asymmetry of Morphology*, which is defined as follows: "Morphology combines and manipulates asymmetric relations only". Her theory is fully worked out for derivation (prefixation and suffixation) and not so much for compounds. Now it would be interesting to explore how her theory can be applied to the compounds analysed in the present thesis, a question that we leave for further research. For now just notice that a first difference between our proposal and Di Sciullo's is that the structures we propose contain a bare sisterhood relation, whereas Di Sciullo's contain "two layers of asymmetric (sister-contain) relations" (Di Sciullo 2005: 35; for a representation see p. 36). Also noteworthy is the parallelism found between our findings and those of Borer (2009) and those of Construction Grammar (cf. Goldberg 1995 and references therein), where compounds can be seen as the concatenation of a non-head which acts as a modifier and a head. Such a comparison will not be pursued here but we hope to take it up in future work.

compounds conform to the structure in (86a) like the adjectival NA compounds, e.g. *cama+curt* (leg+short) 'short-legged' (82a).

The unification of ATR and SUB compounds is not new. While some works on compounding (e.g. Oniga 1992, Val Álvaro 1999) draw a distinction between argumental and modificational compounds, similar to B&S's distinction, other studies (e.g. Pérez Saldanya et al 2004, Scalise 1992) do not, and analyse as 'subordinate compounds' compounds which, in B&S's (2005) view, would be ATR (e.g. *pale face*) and SUB (e.g. *taxi driver*). This second view is similar to our proposal, but notice that each study sharing our view about the non-distinctness of ATR and SUB compounds has CRD compounds as a compound type, whose existence is denied in this thesis.

In short, the novelty of our proposal partially lies in having a unique pattern, i.e. process, of compound creation, which can account for all compound types and from which the different interpretations available arise. The pattern is based on a head vs. non-head relation and the compound denotes a subset of the set of entities denoted by the head. Although we have not said much about how the different interpretations arise, we suggested throughout this chapter that the semantics of the compounding elements - putting special emphasis on the semantic requirements imposed by the head - will determine the final interpretation of the compound (e.g. Pustejovsky 1995, Wisniewski 1996). Our work represents a first step towards *the* (potentially universal) classification of compounds, which at the same time provides support for a morphological analysis of compounding.

Chapter 3. The Morphosyntactic Interface and the Compounding Parameter

In this chapter Snyder's (2001) Compounding Parameter is presented, together with the subsequent amendments it has undergone (section 3.1). The application of the Compounding Parameter to English is summarized in subsection 3.1.1, after which the parameter is applied to Catalan and to other language families (subsections 3.1.2 and 3.1.3). Finally, this first section ends with some discussion about some controversial issues regarding the parameter and suggests some alternatives (subsection 3.1.4).

In the second part of the chapter, two syntactic accounts of resultatives are briefly reviewed: Kratzer's (2005) and Mateu's (2000, 2010). We consider the possibility of extending their analyses to primary compounds (subsections 3.2.1.1 and 3.2.1.2). Next, we address the question of why in languages like Catalan NN compounds are productive although to a lesser degree than NN compounds in English (subsection 3.2.2). Finally, the main findings of the chapter are summarized in section 3.3.

3.1 The Compounding Parameter

This section first presents the basics of Snyder's Compounding Parameter as it was originally proposed. The main source of this presentation is Snyder (2001), from which most examples and quotations are taken, although the same findings are reported in other work (e.g. Snyder 1995, 1996, 2002). Second, some refinements and subsequent revisions added to the original proposal of the Compounding Parameter are briefly discussed (Beck & Snyder 2001a, Snyder et al. 2001, Roeper et al. 2002, Roeper & Snyder 2005, Snyder 2005). Then the Compounding Parameter is considered in English (subsection 3.1.1), in Catalan (subsection 3.1.2) and in other language groups (subsection 3.1.3). Finally, subsection 3.1.4 closes the first part of the chapter with some discussion of controversial questions around the parameter and its alleged implications.

The source of the Compounding Parameter lies in Snyder's (2001) claim that the availability of complex predicates[181] of the type given in (1) is subordinated to the

---

[181] The reasons for choosing the label 'complex predicate' instead of 'small clause' can be found in footnote 21 in chapter 1 and in Snyder (1995: 61, fn. 43).

existence of productive endocentric root compounding (e.g. *frogman*).[182] More specifically, the claim is that a language will only have complex predicates (cf. (1)) if it can form primary compounds productively. That is, there is a strong association between these two types of constructions.

(1)   a. John painted the house red.   (resultative)

           b. Mary picked the book up / picked up the book.   (verb-particle)

           c. Fred made Jeff leave.   (*make*-cause)

           d. Fred saw Jeff leave.   (perceptual report)

           e. Bob put the book on the table.   (*Put*-locative)

           f. Alice sent the letter to Sue.   (*to*-dative)

           g. Alice sent Sue the letter.   (double-object dative)

                                         Snyder (2001: 325)

It seems that the group of complex predicates which are claimed to be dependent on compounding should be enlarged. Beck & Snyder (2001a) argue that telic path(/goal)-PP constructions like *to the summit* in *walk to the summit* should be treated as a type of resultative and Snyder et al. (2001) claim that non-resultative path PPs like *down the banister* in *slide down the banister* should also be included in the group of complex predicates in (1).

       Let us now consider the details of the Compounding Parameter in some depth. From the observation that the complex predicates in (1) are, for example, present in Germanic but absent in Romance, Snyder claims that the availability of such constructions is subject to parametric variation (in the sense of Chomsky 1981).[183] Data from child language acquisition in English seems to corroborate the fact that all these complex predicate constructions form a class which is subject to the same parametric property because of their concurrent acquisition. Snyder claims that the availability of the constructions in (1) hinges on the marked value of a global compounding parameter, which is characterized as follows (Snyder 2001: 328):

---

[182] Recall from chapter 2 that we argued that compounds are all endocentric and that we referred to 'root compounds' as 'primary compounds'. Hence, Snyder's (2001: 328) use of 'endocentric root compounds' will be replaced by 'primary compounds' from now on.

[183] According to Chomsky (1981), parameters are associated with principles of Universal Grammar. Such an approach has been criticized in a number of works, such as Borer (1984) and Chomsky (1993), to which the reader is directed for some difficulties with this approach.

(2)   "THE COMPOUNDING PARAMETER [TCP]: The grammar {disallows*, allows}
       formation of endocentric compounds during the syntactic derivation [*unmarked
       value]."[184]

That is, when the marked value of the parameter is assumed, compounds are derived
syntactically, a fact which Snyder associates with compounding being productive (i.e.
novel compounds can be created spontaneously), which in turn explains the availability
of the complex predicate constructions in (1).[185] This is the case of English.

    If complex predicate formation depends on the availability of syntactic
compounding, it follows that the availability of both constructions should be well
correlated, namely complex predicates should only be available when syntactic
compounding is a possibility. Snyder evaluates this prediction by means of a cross-
linguistic survey in which different language groups are represented. The availability of
syntactic compounding is checked by considering the grammaticality of novel NN
compounds (for details, see Snyder 2001: 330) whereas the availability of complex
predicates is recognised by means of the grammaticality of strong resultatives (for
details, see Snyder 1995: 28-29, Snyder 2001: 330, fn. 10).

    At least two types of resultatives must be distinguished. There is one type of
resultative, which is present in Germanic and totally absent in Romance (3) (cf. e.g.
Levin & Rapoport 1988), and a second type of resultative, which is also present in
Germanic and severely restricted in Romance if it exists at all (4) (see footnotes 186 and
187). The difference between the two types lies in the fact that in the former the
addition of the adjective makes the construction resultative and changes the verb from
being an activity to an accomplishment (i.e. the verb alone is an activity) (cf. Vendler
1967), whereas in the latter the verb already denotes an accomplishment and the
addition of an adjective only makes this fact more evident (although superfluously).
Compare the grammaticality judgments given in (3-4):

---

[184] Snyder's (1995: 27) characterization of the TCP is as follows: "The grammar does (not) freely allow
open-class, non-affixal lexical items to be marked as [+Affixal]." Another formulation of the TCP can
also be found in Roeper et al. (2002), under the label of 'The Root Compounding Parameter', according
to which "Set-merger can(not) combine non-maximal projections" (see also Roeper & Snyder 2005).
[185] Recall that in our view compounding takes places in the morphological component (in the 'word
syntax' in the model of grammar depicted in (11) in chapter 1). As will be seen below, though, the
predictions made by Snyder's (2001) Compounding Parameter can still be corroborated irrespective of the
locus of compound formation.

(3)   a. John hammered the metal flat.

       a'. *En Joan martellejà el metall pla.   (Catalan)

       b. The horses dragged the logs smooth.

       b'. *Els cavalls arrossegaren els troncs llisos.   (Catalan)

(4)   a. John painted the house red.

       a'. En Joan pintà la casa vermella.[186]   (Catalan)

       b. Mary froze it hard.

       b'. *La Maria ho congelà dur.   (Catalan)

These two types of resultatives have been called strong and weak resultatives in Washio (1997) where they are defined in the following terms. Strong resultatives are those in which the information provided by the adjectival phrase is not predictable from the lexical semantics of the verb, as in the examples illustrated in (3). For example, in (3a) as a consequence of the hammering process, the metal can become flat but it can also become shiny, soft, etc., states that are not implied by the meaning of the verb. By contrast, in weak resultatives, the verb implies a state that the patient might come to be

---

[186] Not all Catalan speakers agree on the acceptability of (4a'). Consider other examples of resultatives available in Romance.

  (i)   El pare fregà la taula   ben neta.   (Catalan)
           The father wiped the table-FEM.SG very clean-FEM.SG
  (ii)   El helado   se   congeló bien congelado.   (Spanish)
           The ice.cream-MASC.SG CL.REFL froze well frozen-MASC.SG
  (iii)   Ho stirato la camicia piatta *(piatta).   (Italian)
           I.have ironed the shirt-FEM.SG flat-FEM.SG (*flat-FEM.SG)

One common feature of the resultatives in (i-iii) is the use of devices to emphasize the result predicate: note the use of *ben* 'very' in (i), *bien* 'well, very' in (ii) (example from Demonte & Masullo 1999: 2470; see also Demonte 1992) or the doubling of the adjective *flat* in (iii) for the meaning of 'very flat' (example from Napoli 1992: 74-75, ex. 109b, 112). These emphatic devices are not necessary in languages which have similar resultatives. Consider English in (4a, b) or Japanese in (iv) (example from Washio 1997: 10, ex. 29):

  (iv)   Mary-ga doresu-o pinku-ni some-ta.
           Mary-NOM dress-ACC pink   dye- PST
           'Mary dyed the dress pink'

Other resultative constructions which should be differentiated from those in (3) are those which include a light verb and an adjectival predicate. This construction is present in both English and Romance. Consider the following example from Catalan (on this point, see Mateu 2002, Rigau 2002).

  (v)   El ferrer   deixà el metall pla.
           the blacksmith left   the metal flat
           'The blacksmith flattened the metal'

in as a result of the action named by the verb, as in (4). For instance, the verb *paint* in (4a) encodes the notion colour and the adjective *red* specifies which colour it is. That is, the verb *paint* has a 'disposition' towards a certain state, that of being painted in a certain colour.[187] (On the distinction between weak and strong resultatives, see also Kaufmann & Wunderlich 1998).

Variation in judgment and the use of emphatic devices in the resultatives available in Romance languages (see footnote 186) makes it difficult to determine whether such resultatives should fall into the group of weak resultatives or should rather be treated as a different phenomenon. For our present purposes, the choice is irrelevant since the resultative construction relevant to Snyder's parameter is the strong one. In what follows, a distinction between strong resultatives (those depicted in (3)) and non-strong resultatives (including weak resultatives proper and the resultatives present in Romance (4), see footnotes 186 and 187) will be made and the terms 'resultatives' and 'strong resultatives' will be used interchangeably to refer to the resultatives relevant in Snyder's survey. In the case confusion may arise the terms 'strong', 'non-strong' or 'weak' will be explicitly used.

In short, it is the (un)grammaticality of the strong resultative (3) which must be used as a diagnosis of the (un)availability of the complex predicates (1) in Snyder's survey, for the alleged correlation to work (see below).

The following table reproduces the findings of the survey (borrowed from Snyder 2001: 329).

---

[187] Washio (1997: 17) also talks about a third type of resultative: spurious resultatives, although he convincingly argues that they are not resultative expressions. The paraphrase generally accepted for resultatives "x causes y to become z" often fails with spurious resultatives; the adjectival predicate describes the manner in which the activity named by the verb is carried out and adjectives can alternate with adverbs with no change in meaning (for other properties, the reader is referred to the original work). (see also Levinson 2010). Consider the following examples:

(i)   He tied his shoelaces tight / tightly (≠ He caused his shoelaces to become tight by tying them)
(ii)   He spread the butter thin / thinly (≠ He caused the butter to become thin by spreading it)
(iii)   He cut the meat thick / thickly (≠ He caused the meat to become thick by cutting it)

Spurious resultatives, also available in Romance, are irrelevant to Snyder's parameter.

(iv)   Talla-les   menudes.   (Catalan)
        Cut+them.FEM fine-FEM.PL
(v)   Mia figlia   ha cucito la gonna (troppo) stretta.   (Italian, Washio 1997: 30, ex. 90)
        My daughter has sewed the skirt   (too)   tight

Table 3.1: Cross-linguistic survey of resultatives and NN compounding

	RESULTATIVES	PRODUCTIVE N-N COMPOUNDING
American Sign Language (ASL)	yes	yes
Austroasiatic (Khmer)	yes	yes
Finno-Ugric (Hungarian)	yes	yes
Germanic (English, German)	yes	yes
Japanese-Korean (Japanese, Korean)	yes	yes
Sino-Tibetan (Mandarin)	yes	yes
Tai (Thai)	yes	yes
Basque	no	yes
Afroasiatic (Egyptian Arabic, Hebrew)	no	no (?)
Austronesian (Javanese)	no	no
Bantu (Lingala)	no	no
Romance (French, Spanish)	no	no
Slavic (Russian, Serbo-Croatian)	no	no

The table shows that there is a strong correlation between the two types of constructions (i.e. resultatives and NN compounds) and that the relationship is directional. Basque has compounds and yet has no resultatives, which suggests that compounding is seemingly a necessary but not a sufficient ingredient for the availability of resultatives (and complex predicates more generally).[188]

The question of how compounding and complex predicate formation are connected is given different answers in Snyder's work. Initially the dependence relation of complex predicates on compounding is based on the claim that complex predicates "involve a *morphological compound* at some abstract level of grammatical representation, (…)" (p. 328), a level which is identified with "the point of semantic interpretation (LF)" (p. 336). The connection between the two types of constructions is then semantic. Snyder argues for a type of semantic composition, which is available in

---

[188] The same generalization can be reached by looking at separable-particle constructions, like the English *Mary **lifted** the box **up*** (see, for example, the cross-linguistic survey in Snyder 2002).

compounding and of which complex predicate constructions also make use. The mode of semantic composition is summarized in the following constraint (p. 336):

(5)   "COMPLEX-PREDICATE CONSTRAINT: Two syntactically independent expressions can jointly characterize the event-type of a single event-argument, only if they constitute a single word (endocentric compound) at the point of semantic interpretation."

Accordingly, (3a) can describe an accomplishment because *hammer* and *flat* together form a compound at LF. One could say that the verb combines with the secondary predicate in such a way that the combination behaves as a simple verb semantically. In other words, the predicates *hammer* and *flat* are viewed as a single predicate (a complex predicate) which takes the argument *the metal*. That this seems to be the case is shown by the contrast in grammaticality of the following sentences.

(6)   a. John hammered the metal (for an hour)/(??in an hour).
       b. John hammered the metal flat (?for an hour)/(in an hour).

<div align="right">Snyder (2001: 326, ex. 2a, d)</div>

The addition of the durational modifier *for an hour* is only allowed with the activity verb *hammer*. The same durational modifier is not fully accepted when *hammer* combines with *flat* (hence, the question mark), which suggests that an accomplishment has been created as a consequence of the verb *hammer* forming part of the complex predicate *hammer flat*. As a result, only the aspectual modifier *in an hour* is fully accepted.[189] Bear in mind that the view presented above is different from Snyder's (1995: 45-59), where one phonologically null aspectual morpheme (labelled ϕ_{telic}) mediates between the two visible elements that form the resultative (i.e. *hammer* and *flat*), and in fact the relevant compound to be interpreted at LF is formed by *hammer* and the null aspectual morpheme, with *flat* being a restrictor on a subpart of the event characterized by the compound. What Snyder (1995) and Snyder (2001) have in common is that complex predicates participate in the creation of a 'complex word' (be it

---

[189] This view is in agreement with a Montagovian approach to semantic composition, according to which syntactic positions and arguments of a predicate are mainly in a one-to-one relationship (Dowty et al. 1981).

directly or indirectly) at some point in the syntactic derivation, and that this is only possible in languages which take the positive setting of the TCP in (2).

Although in Snyder (2001) the connection between primary compounds and complex predicates is established by sharing the mode of semantic composition given in (5), it is not clear how (5) applies to compounds if they never define the event-type of a single event-argument, unlike complex predicates.

In Beck & Snyder (2001a) the connection between compounding and complex predicate formation is not provided. It is argued that telic path PPs like *walk **to the summit*** are assimilated into resultatives (*hammer the metal flat*) and that constructions like these, together with the verb-NP-particle construction like in *lift the box up*, are allowed only in [+TCP] languages where a rule of semantic composition called Principle R applies. Such a principle includes as semantic primitives CAUSE and BECOME and is responsible for yielding a resultative reading by combining non-resultative predicates: *John walked to the summit (in an hour)* is given the paraphrase "John's walking caused him to become at the summit" (p. 117). Beck & Snyder (2001a: 116) make the following statement:

(7)   "When root compounding [primary compounds in our terms] is available as a mechanism of syntactic combination, syntactic sisters can freely be treated as forming a complex word, for purposes of semantic interpretation."

Accordingly complex predicates are seen as a complex word in [+TCP] languages (much as in Snyder 2001) and given that in Beck & Snyder's understanding Principle R is available only within a complex word (p. 116), such a principle can apply to complex predicate constructions and derive the expected (resultative) reading. However, Principle R cannot be available in all [+TCP] languages, because there are languages like Basque which have productive primary compounds and yet complex predicates are absent. Beck & Snyder are then forced to propose that Principle R is subject to parametric variation and that languages like Basque (and Catalan, see below) lack Principle R: "We thus propose that Basque has productive root compounding, but lacks Principle R." (p. 116). This solution shows that Principle R is irrelevant to compounding, which suggests that the availability of complex predicates is dissociated from the availability of primary compounds. We will propose in sections 3.1.4 and 3.2 that complex predicates are not dependent on the availability of productive

compounding and that the two constructions (complex predicates and compounds) are constrained by different factors. If complex predicates like resultatives and telic path PPs required the operation of compounding, as Snyder (1995 and subsequent work)[190] claims, then a logical possibility would be that both complex predicates and compounds were subject to the same modes of semantic composition, Principle R being one of them. Despite primary compounds forming a complex word (a prerequisite for Principle R to operate), Principle R must be prevented from applying to them because they do not have a resultative interpretation which the Principle R is designed to derive.[191] On this speculative note, it would not be clear why complex predicates and compounds behave differently with respect to the so-called Principle R. In short, because Principle R does not apply to compounding, the connection between compounding and complex predicates is left open in Beck & Snyder (2001a).

Principle R is revised in Snyder et al. (2001): the semantic primitive BECOME is deleted from Principle R in order to account for non-resultative path PPs like *down the banister* in *slide down the banister*. To explain the BECOME component in constructions like adjectival resultatives (*hammer the metal flat*) and resultative path PPs (*walk to the summit*), the former is claimed to include a null morpheme BECOME and in the latter the preposition is analysed as BECOME AT (e.g. the preposition *to* in *walk to the summit*). Despite the amendment made to Principle R, its revised version still does not help explain how compounding and complex predicates are connected: primary compounds are unaffected by Principle R. Let us now turn to Snyder's (2005) most recent proposal regarding the relation of compounding to complex predicate formation.

In Snyder (2005) Principle R is replaced by another semantic composition rule, called Rule C, which is also subject to parametric variation but, unlike Principle R, Rule C is required both for the interpretation of novel compounds and for the formation of complex events like accomplishments (out of simple event predicates). As for the building of complex events, the proposal is similar to the one contained in (5) in the sense that it is assumed that verbs take a Davidsonian event argument and that the second predicate (e.g. a path PP) also takes an event argument (Davidson 1967). The event arguments of the two predicates are identified to characterize a single, complex

---

[190] Given that the reference of 'Snyder (1995 and subsequent work)' is used in a very high frequency in the present chapter, it will be shortened to 'Snyder (1995f)' to avoid clumsiness.
[191] Although the semantic primitive CAUSE could be argued to exist in compounds like *drug deaths* (Levi 1978, see section 2.2.2 in chapter 2), it is difficult to maintain such a position for many compounds (e.g. *soldier ant*, *apple pie*, *roads lobby*, *anteater*, *car mechanic*). The same difficulty arises with the semantic primitive BECOME.

event, a process which is carried out by means of Rule C, which is given the following characterization (p. 3):

(8)   "If a = [b c], and b' and c' both have an open argument position of semantic type x, then (ignoring any other open argument positions) a' = c' OF THE KIND ASSOCIATED WITH b'."

"Rule C can apply to predicates of events <u>or</u> predicates of individuals"

Application of Rule C to predicates of events is designed to account for the interpretation of complex predicates, like the one formed by a verb plus a resultative path PP, as in *The bottle floated under the bridge*, which is given the following interpretation: "There exists a (past) event of the bottle floating, and this event is <u>of the kind associated with</u> the bottle moving to a location under the bridge". Application of Rule C to predicates of individuals explains how the interpretation of primary compounds is obtained. For example, *frogman* is given the interpretation of being a "man <u>of the kind associated with</u> frogs".

On Snyder's account, the availability of Rule C implies that primary compounds are productive ([+TCP] languages) and have compositional semantics. Within this new approach, the TCP is revised as follows:

(9)   "Rule C {is, is not} available at the syntax/semantics interface."

In short, the connection between compounding and complex predicates is explicit in Snyder (2005). Both constructions are interpreted by means of the same semantic composition mode: Rule C, which suggests that if such a rule is available in a language, both compounding and complex predicates should also be available. That these constructions are not interdependent is shown by languages like Basque and a few other languages (see below). On this account, complex predicates do not form a compound at some point in the syntactic derivation (unlike in Snyder 2001), although they still imply the availability of productive compounds in the language.

After having presented the basics of the TCP and the alleged dependence relation of resultatives on the availability of primary compounds (NN compounding), let us now summarize how the TCP and the alleged correlation fare in English (subsection

3.1.1). The TCP and the alleged correlation will also be considered in Catalan (subsection 3.1.2) and in other language groups (subsection 3.1.3). The findings of Snyder's survey (Table 3.1) are used as the starting point for our examination in the following subsections. Finally, some controversial issues concerning the TCP follow (subsection 3.1.4).

3.1.1 English

From the discussion above it is clear that the TCP is set to the marked value in English because primary compounds are productive. Recall from subsection 2.3.1.1 in chapter 2 that [NN]$_N$ compounds are the most productive type of compounding in English (see the data in (32), (35) and (36)), which would in turn explain why strong resultatives (and more generally the complex predicates in (1)) are available. The English data fit well with the putative correlation between productive primary compounds and the availability of resultatives.

3.1.2 Catalan

According to the results of the survey in Table 3.1, Romance languages like French and Spanish have neither strong resultatives nor productive primary compounds. These facts can be accounted for by appealing to the unmarked setting of the TCP: the absence of productive primary compounds explains the unavailability of the strong resultative.

However, we want to argue against the claim that Romance languages have no productive primary compounds. Recall that the conclusion from chapter 2 when dealing with nominal compounds in Catalan (subsection 2.3.2.1) was that NN compounds are productive in the language (see the discussion around the data in (64) and (65b-c)). Although one could argue that compounds like *faldilla pantaló* (skirt trousers) 'skort' and *verd oliva* (green olive) 'olive-green' (both included within the examples in (64)) are in some sense lexicalized, one can easily create novel NN compounds: *un jardí museu* (a garden museum) 'a garden which is also a museum', *una maleta maletí* (a backpack briefcase) 'a backpack which resembles a briefcase' and *una piscina aquari* (a swimming.pool aquarium) 'a swimming pool which may have fish like an aquarium'.[192]

---

[192] Interestingly, acquisition data on compounding reveal that NN compounds are a common interlanguage strategy among adult L2 learners of Spanish, including French speakers. According to Snyder (1995f), both Spanish and French have the TCP set to the negative value (like Catalan) and yet both attested and non-attested NN compounds were produced by French speakers when labelling some pictures (with real or fictitional entities) shown to them (on this point, see Liceras et al. 2002).

In short, Catalan would be like Basque in the sense that they both have NN compounding but lack strong resultatives. This conclusion requires Table 3.1 be revised, but does not deny the alleged dependence of resultatives on NN compounding. Snyder suggests that, in addition to the availability of NN compounding, other prerequisites may be necessary for strong resultatives to be available. In the case at hand, we must then conclude that the availability of NN compounding in Catalan is not sufficient and that another factor (or other factors), which is lacking in the language, is necessary.

3.1.3 Other language families

Snyder (1995: 31) presents a table slightly different from Table 3.1 (borrowed from Snyder 2001: 329). The differences have to do with the placement of some languages, namely ASL, Japanese and Mandarin. Whereas they are treated as languages with resultatives (e.g. *John hammered the metal flat*) and productive NN compounding (e.g. *worm can*) in Snyder (2001 and subsequent work), they are treated as not having such constructions in Snyder (1995). Such a divergence is accounted for by the use of more flexible criteria in Snyder's more recent work: "In Snyder 1995 a potential resultative construction was excluded if it contained any material absent from the English resultative, such as the ASL word glossed as BECOME (…). In the present study, the element BECOME in ASL, and *haj* in Thai, are regarded as possible overt counterparts to a null morpheme in the English resultative (…)" (Snyder 2001: 330, fn. 10; see also Beck & Snyder 2001a: 120, fn. 2). This explains why ASL is claimed to have resultatives in Table 3.1. According to Snyder, if one language has resultatives, it must necessarily have productive compounding, a correlation which seems to be present in the case of ASL (see Table 3.1).[193] Such a correlation, though, seems to be questioned in Snyder (1995: 32, fn. 6), where it is said that "N-N compounding in ASL is thus distinguished from that in English both by a relative lack of productivity and (…)". From these contradicting results regarding the status of compounding, it is difficult to determine whether the putative compounding/resultative correlation holds for ASL.

---

[193] Snyder (2001: 338) provides the following example, which seems to point to the availability of NN compounding in ASL.

(i)   BANANA BOX (for 'a box in which bananas are stored')

Concerning Japanese, it is claimed to have both productive compounding and strong resultatives in Snyder (2001 and subsequent work). However, Snyder (1995: 32, fn. 10; 65) makes the following statements:

(10) "Despite the existence of lexical N-N compounds in Japanese, my informants judge novel N-N compounds (as for "worm can") to be possible only as an attempt at lexical innovation; where English would freely permit the spontaneous creation of a novel N-N compound, Japanese normally requires a phrasal construction with the connector *no*." and "(…) Japanese (…) lacks productive N-N compounding (…)"

In other words, if Japanese really has resultatives but has no productive NN compounding, Snyder's claim that resultatives are dependent on the availability of productive compounding is falsified. Snyder's methodology used to test resultatives in Japanese, though, is questionable: children passed the resultative task (a truth-value judgment task) if they answered "correctly on all three resultative/attributive examples with *nuru* 'paint', or with *kiru* 'cut', or both." The examples with *paint* are illustrated below (Snyder 2002: 37-38):

(11) a. Pikachu-wa aka-i isu-o nutte-imasu.
        'Pikachu is painting the red chair.' (attributive example)
      b. Pikachu-wa aka-ku isu-o nutteiru.
        'Pikachu is painting the chair red.' (resultative example)

Although this study concludes that resultatives are a possibility in Japanese, and so does Snyder (2001), notice that the resultative in (11b) is a non-strong resultative and so is the resultative used in Snyder (2001: 337), namely *to wipe the table clean*.[194] The presence of non-strong resultatives in Japanese does not question Snyder's compounding/complex-predicate parameter, since the parameter is sensitive to strong resultatives and these are absent (see Washio 1997, Tomioka 2004).

In addition, Snyder's (1995) claim that Japanese has no productive primary compounds can also be questioned. The availability of productive compounding was

---

[194] See Washio (1997: 12-16) for discussion of *to wipe the table clean* as a weak resultative.

tested by giving informants a context in which they had to judge the direct counterpart of *worm can* in their language. In Japanese the word preferred for *can* is strongly associated with foodstuffs, which explains why the test for primary compounds gave a negative result in Snyder (1995). The test was changed in his more recent work and the result is that Japanese does seem to have productive primary compounds (see, e.g., Beck & Snyder 2001a: 120, fn.1).[195] It seems then that Japanese is like Basque and Catalan in that it has primary compounds but no strong resultatives. In short, Snyder's implicational relationship from resultatives to productive primary compounds can still hold.

Regarding Mandarin, the availability of resultatives was tested in Snyder (1995) by means of a weak resultative (*to paint the house red*) with a negative result. Snyder (2001), nonetheless, used the strong type of resultative (*to beat the iron pipe flat*) to test the availability of resultatives, which gave a positive result. As for the availability of compounding in Mandarin, the change from its unavailability to its availability seems unwarranted, conveniently made to fit Snyder's prediction: if Mandarin has strong resultatives, it should also have productive compounding. The data below show mixed results: the data in (12b) seem to point to the fact that Mandarin has nominal compounding of the English type (e.g. *worm can*), but the connecting device *de* in (12a) seems to question it.

(12)   a. zhuang   chong de   guan (Snyder 1995: 34; tones omitted)
          store/put worm DE can (for "worm can")
       b. you ji (Li & Thompson 1981: 50; cited in Snyder 2001: 338; tones omitted)
          oil   stain

In short, the changes made in the cross-linguistic survey depicted in Table 3.1 with respect to the data of the table illustrated in Snyder (1995: 31) cannot all be accounted for by the use of more flexible criteria. Some changes are not given an explanation, which makes them unwarranted. In the next subsection it will be seen that the validity of Snyder's compounding/complex-predicate parameter is further weakened.

---

[195] Snyder (2001: 338) provides the following example.

(i)   bananabako
         banana + box

3.1.4 Discussion

After having seen that the alleged correlation between NN compounding and strong resultatives is not as strong as is claimed by Snyder (1995f), other controversial points regarding the TCP will be considered. Some are minor points but other questions really threaten the implicational relationship from complex predicates to the positive setting of the TCP as well as the interdependence among the complex predicates, which are argued to form a natural class by Snyder.

First, it is not clear why the marked value (as opposed to the unmarked one) of the TCP is responsible for the availability of productive primary compounds, which in turn explains the availability of the complex predicate constructions in (1). That is, the notion of '(un)marked value' is problematic in the sense that it is not obvious on what basis one decides which is the (un)marked value for the parameter. In relation to this, Snyder (1995: 27, fn. 2) provides no satisfactory answer: "(…) undoubtedly as the result of deeply ingrained anglocentrism, I have persisted in stating (6) [TCP] so that English, rather than French, receives the positive setting of the parameter." From this quote it may seem that as long as complex predicates and productive compounding rely on the same value of the parameter, be it marked or unmarked, Snyder's analysis can go through. Liceras et al.'s (2002: 229-230) study of acquisition data on compounding, though, provides evidence against NN compounding being the result of the marked option of the TCP: since interlanguage speakers refrain themselves from producing marked constructions and yet NN compounding is a "very productive interlanguage strategy", it seems that NN compounding cannot be the result of the marked option of the TCP.

Second, another unclear point is why novel NN compounds are chosen as the diagnosis of productive compounding and complex predicate formation.[196] Despite NN compounding being the most common compound type in English, it is not the only type of compound which satisfies the requirement of being a productive compound and does

---

[196] Some clarification remarks are in order here. Snyder (1995: 27) does actually mention other types of primary compounds, namely [AN]$_N$ and [VN]$_N$ compounds (e.g. *blackbird* and *guard dog* respectively), as other possible compound types which can be used as a diagnostic for complex predicate formation. However, the [AN]$_N$ and [VN]$_N$ compound types are not productive in English, which explains why Snyder (2001) refrains from mentioning such compounding types as potential diagnoses of complex predicates.

not mean that the same type of compound is also the most frequently used in the other language groups included in the survey. Recall from chapter 2 that other productive compound types in English include, among others, the verbal type [NV]$_V$ (e.g. *computer-generate* (46a)) and the adjectival type [NA]$_A$ (e.g. *oil-rich* (53d)). As for other languages included in Snyder's survey, Table 3.1 shows that Romance languages lack productive NN compounds. We argued against this conclusion. Earlier in this chapter as well as in chapter 2 we showed that Catalan has productive NN compounding and that, in addition, Catalan possesses other productive compounds, some of which follow: the nominal types [VN]$_N$ (e.g. *busca-raons* (look.for+reasons) 'troublemaker' (68)) and [NA]$_N$ (e.g. *relacions catalano-occitanes* (relations Catalan+Occitan) (83b)) and the adjectival type [NA]$_A$ (e.g. *cama+curt* (leg+short) 'short-legged' (82a)). In short, it is not obvious why NN compounding, as opposed to other types of compounds, is used as a diagnostic for productive compounding and why it should be a prerequisite for the availability of complex predicates.

Third, a question related to the previous one is how compounding relates to complex predicates, a matter which has been given different accounts in Snyder's work but none of them seems to be satisfactory (see the discussion in section 3.1). Snyder (2001: 336) recognises that more research into the connection between compounding and complex predicate formation is needed. To this end we will now consider the possibility that the connection has to do with the categories involved.

If one takes Snyder's claim seriously, namely that complex predicates and compounds are strongly associated and that the existence of the former depends on the availability of productive primary compounds (without requiring they be nominal NN compounds), one may expect that the categories involved in a complex predicate should also be present in a compound type. If this were the case, it would indicate that compounding is really a prerequisite for complex predicate formation, as Snyder argues. To establish whether this correlation does or does not hold, English will be used: we will examine whether the categories involved in the complex predicates in (1), repeated below for convenience, are also present in some compound type. The conclusion will be that the alleged correlation is questioned: the categories involved in the complex

predicates are not present in compound types, or if they are the compound is not productive. Each complex predicate in (1) will be considered in turn.[197]

(1)   a. John hammered the metal flat.   (resultative)

           b. Mary picked the book up / picked up the book.   (verb-particle)

           c. Fred made Jeff leave.   (*make*-cause)

           d. Fred saw Jeff leave.   (perceptual report)

           e. Bob put the book on the table.   (*Put*-locative)

           f. Alice sent the letter to Sue.   (*to*-dative)

           g. Alice sent Sue the letter.   (double-object dative)

                                    Snyder (2001: 325)

If the English resultative construction (1a) involves a verb and an AP (e.g. [*to hammer$_V$ flat$_A$*]$_{VP}$), the same two categories (i.e. verbs and adjectives) are predicted to merge in a compound. Accordingly, two compound types are predicted to exist: the [VA]$_A$ compound and the [AV]$_V$ compound. As for the former, it is nonexistent in the language. This conclusion was reached in chapter 2 (subsection 2.3.1.3) when considering the scarcity of such forms in the language (e.g. *diehard*, *fail-safe* (54)) and the speaker's inability to create new forms based on this pattern, which is in agreement with the findings in Booij (2005), Plag (2003) and Selkirk (1982) (contra Carstairs-McCarthy 2002). As for the latter (i.e. the [AV]$_V$ compound), we concluded in chapter 2 (subsection 2.3.1.2) that it is a very limited compounding process, despite giving the same treatment to all AV compounds: base-generated and derived compounds (51). Recall that although some AV compounds are often claimed to be derived from nominal or adjectival compounds via back-formation (e.g. *literary-editor > literary-edit*) or conversion (*blackmail*) (cf. Bauer 1983, Plag 2003) and, in fact, all AV compounds may have a nominal/adjectival counterpart (*free associate ~ free association*), AV compounding is a possibility in the language, although limited. Otherwise, nominal and adjectival compounds would not undergo back-formation and conversion to a verbal AV compound. The lack of the [VA]$_A$ compound and the limited [AV]$_V$ compounding process predicts that the combination of verbs and adjectives (APs) in syntax will be

---

[197] Recall that the example with a weak resultative in (1a) (*John painted the house red*) has been replaced by the sentence in (3a) (*John hammered the metal flat*), which contains a strong resultative, in order to test the validity of the compounding/complex-predicate parameter, which is sensitive to strong resultatives only.

either nonexistent or very low in productivity. The reality, though, points in the opposite direction. Resultative constructions of the type illustrated in (1a) are common in English, hence allowing the merger of verbs and adjectives productively, contra our prediction.

The constituting elements of the complex predicate in (1b) are a verb and a particle (P) (*pick up*), which seem to be the same elements present in the complex predicates in (1e, f): *put on* in (1e) and *sent to* in (1f).[198] Given that verbs merge with Ps productively in syntax giving rise to complex predicates of the type illustrated in (1), the prediction is that the V+P/P+V combination must also be present in some compound type (if compounding is really a prerequisite for complex predicates). In chapter 2 we described four different formations which include the V+P/P+V merger but we argued that none of them is a compound type. Let us consider each in turn. Whereas two of the formations result in a noun, the other two formations result in a verb. The two nominal forms are [VP]$_N$ and [PV]$_N$. We showed that examples of the [VP]$_N$ type (cf. (42)) are either cases of suffixation (e.g. *V+in*: *laugh-in*, *love-in*) or cases of converted syntactic phrases, namely phrasal verbs converted to nouns (e.g. *breakdown*) (see Berg 1998). The conclusion is that there is no [VP]$_N$ compounding process. Nor is there a [PV]$_N$ compounding process: forms like *downfall* also come from syntactic constructions converted to nouns (cf. (45)). As for the verbal forms also mentioned in chapter 2, one of them is phrasal verbs, the sum of V+P, which we take as a syntactic object and not as a compound, hence their omission from the compound types available in English. As for the other verbal formation, the sum of a P+V, it is not a case of compounding either. Whereas some [PV]$_V$ forms are instances of prefixation (e.g. *out-* in *outrun*, *outswim*), others are related to phrasal verbs (e.g. *downplay ~ play down*) (Adams 2001). In short, the lack of PV/VP compounding cannot explain why such categories can merge in syntax in a productive way, an unexpected correlation under the assumption that the categories involved in compounding define which complex predicates can be available.

The complex predicates involved in (1c, d) are two verbs: *made leave* and *saw leave*. Given that two verbs can merge syntactically giving rise to a complex predicate, one would expect a sequence of two verbs to be also present in compounding. Although verbal compounds of the [VV]$_V$ type, like [AV]$_V$ compounds, may all have a nominal or

---

[198] We follow authors like Svenonius (1996) and Stiebels (1998) in treating particles as prepositions. The two objects are referred to as P in the text. A reductionist view is also endorsed by other authors like Ackerman & LeSourd (1997:1), who take the term 'preverb' to subsume prefixes, proclitic elements and particles.

adjectival counterpart (e.g. *crash-landing ~ crash-land*), from which they could be argued to be derived with the result that one could hold that there are no VV compounds, we maintained in chapter 2 (subsection 2.3.1.2) that VV compounding does exist as a compounding process in the language (cf. Booij 2005, Plag 2003; contra Selkirk 1982). Otherwise, nominal and adjectival compounds would not result in VV compounds, if these were not allowed by the grammar. However, the number of attested VV compounds is small (see the examples in (49a, b) of chapter 2: *dive-bomb*, *freeze-dry*). The limited number of VV compounds and the difficulty in creating novel instances of such a compounding type leads us to predict that the merger of two verbs in syntax should not be common. Such a prediction is not corroborated by the data: the syntactic constructions in (1c, d), which involve the merger of two verbs, are not unusual in the language, to say the least.

Finally, the complex predicate in (1g) consists of a verb and a DP: *sent* and *the letter*. Functional material like determiners are excluded from appearing in productive compounding (see chapter 2), which means that the categories present in the complex predicate are not part of any compound type. To rescue the alleged correlation between complex predicate formation and compounding, one could argue that the categories involved in (1g) are a verb and an abstract preposition, similar to the categories physically visible in (1f). The option of merging verbs and prepositions in compounding, though, has already been discarded above (see the discussion of (1b, e, f)). One could also argue that the counterpart of a DP in a compound is a noun with the result that nouns and verbs are expected to merge in compounding. Two compound types with the merger of verbs and nouns were identified in chapter 2: $[VN]_N$ and $[NV]_V$. As for the former, it is very limited in productivity and most compounds of this type are lexicalized. This conclusion was reached in chapter 2 (subsection 2.3.1.1) when compounds like *search party* (37a) and *pickpocket* (37b) were considered. As for the latter, even though $[NV]_V$ compounds (like AV and VV compounding) might be argued to be derived (e.g. *proof-reading* or *proof-reader* > *proof-read*, *handcuff*$_N$ → $_V$), we concluded in chapter 2 (subsection 2.3.1.2) that they are a compound type available in the language and that it is productive (e.g. *computer-generate* (46a), *chain-smoke* (46b)). In short, the existence of the complex predicate in (1g) might be predicted from the availability of verbal NV compounding. However, this putative correlation is the only one found among all complex predicates in (1), which makes it look somewhat coincidental.

In conclusion, there are empirical problems if lexical categories define Snyder's alleged correlation between complex predicates and compounds. The complex predicates in (1), with the exception of (1g), do not have corresponding productive compound types with the same categories. In other words, Snyder's claim that complex predicate formation depends on compounding cannot be established on the basis of the categories involved in the two constructions. The link between the two constructions remains unclear, a point which will be further illustrated in the next question to be discussed.

Last but not least, a question related to the first point discussed in the present subsection is why some languages which seemingly take the unmarked value of the TCP have some of the complex predicates in (1). If all complex predicates in (1) are learned as a group, as Snyder (2001) claims, and their availability depends on the marked value of the TCP, the prediction is that languages taking the unmarked value of the parameter should lack all types of complex predicates. Why is it the case then that languages like Catalan which are supposed to assume the unmarked value of the parameter have some of the complex predicates at their disposal (e.g. *put*-locative, *to*-dative, double-object dative, see below). Snyder (1996: 734, fn. 1; 2001: 326, fn. 3) is aware of this problem and in a footnote he suggests that apparent complex predicates in Romance are only superficial counterparts to English complex predicates. The solution given by Snyder is that the forms in (1) can be structurally ambiguous. Similarly but in the opposite direction, languages which are supposed to have all the complex predicates at their disposal - because the marked value of the parameter is assumed - may lack some of them: "the Germanic languages, which generally resemble English in permitting *most* of the constructions in 1, do not necessarily permit ALL of the constructions." (Snyder 2001: 326, fn. 3).

As just noted, Snyder (1995f) assumes the [-TCP] value for Romance languages. Although it is true that the ease with which NN compounds are created in English is not present in Catalan (a point to which we will return in subsection 3.2.2), we have already shown that novel instances of NN compounds are possible in Catalan (see for example chapter 2 and subsection 3.1.2 of the present chapter for discussion and examples). Accordingly, the marked value of the TCP should be taken, which is in contradiction with the fact that Catalan does not allow strong resultatives (cf. section 3.1, footnotes 186 and 187) or verb-particle constructions (but see below). Consider the

Catalan examples in (13a', 13b'), which correspond to the Spanish examples in (13a, 13b) provided by Beck & Snyder (2001a: 115):

(13) a. María golpeó el metal (*liso).
      a'. La Maria colpejà el metall (*llis).
         The Mary beat   the metal   flat
        'Mary beat the metal flat'.
      b. Chris levantó (*alto) el libro.[199]
      b'. En Chris aixecà (*alt) el   llibre.
          The Chris lifted   up the book
        'Chris lifted up the book'.

Regarding the absence of verb-particle constructions in Romance, some qualifications are in order. Romance languages like Italian, Catalan and Spanish do have verb-particle constructions, but they are of a different nature from those available in English (see e.g. Masini 2005, Iacobini and Masini 2007, Mateu & Rigau 2009, 2010). Mateu & Rigau (2009, 2010) show that in Romance the verb always encodes or involves a directional (path) meaning, which is further specified by a particle (e.g. Italian: *uscire fuori*, lit. 'exit out', *correre via* 'run away'). Such a directional requirement on the verb is not present in the English verb-particle construction (where the direction is expressed by means of the particle). Consider *Gianni danced away*, whose Italian counterpart is impossible: **Gianni* é *danzato via* (Gianni is danced away). From this we conclude that the verb-particle construction relevant to the [TCP] must be the one available in English.

When faced with languages like Basque (and Catalan), Snyder recognizes that taking productive, primary compounds as the relevant type of compounding in the TCP cannot be the only prerequisite for having complex predicates (1), as he intended it to be initially (Snyder 1995: 25, 29, 53, 155). In Snyder's (2001: 330) terms, "root compounding [primary compounds using our terminology from chapter 2] is a necessary, but not sufficient condition for the availability of resultatives." Whenever NN compounding is available in a language and complex predicates like strong resultatives are not a possibility, Snyder appeals to an easy solution to get the facts

---

[199] The sentence would be grammatical if *alto* were interpreted as an adverb (see Bartra & Suñer 1992).

right: he argues that, in addition to NN compounding, some additional unknown factors are necessary.

In front of this panorama, one might try to argue that primary compounds are still a good predictor of complex predicates if they involve the right semantics. That is, one could argue that the semantics involved in the relevant English compounding is different from the one found in its Catalan counterpart and that it is only the semantics involved in English compounding that matters for the availability of complex predicates. For example, the interpretation of the novel English compound *worm can* is that of "a can in which fishing bait is stored" (Snyder 2001: 338), an interpretation which does not seem to be available to the Catalan counterpart *pot cuc* (can worm) which can be interpreted as a can which has the shape of a worm, a can which can be changed into a worm, etc. In other words, the goal/locative reading of the English compound seems difficult to obtain. Although some specific semantics might be required in primary compounds for them to be a good predictor of complex predicates, such a possibility has to be discarded since it would be able to explain the facts only partially. It is true that the direct counterparts to strong resultatives and separable-particle constructions available in English are ungrammatical in Romance (13), which could be explained by appealing to the fact that NN compounding does not have the required semantics for complex predicates to be permitted, but it is also true that other complex predicate constructions are present in languages like Catalan, a fact to which we now turn.

That is, we want to argue against Snyder's (1996: 729, 2001: 326) claim that Romance languages lack all the complex predicates listed in (1)[200] and that those complex predicates which seemingly exist in Romance are not examples of complex predicates, as they are in English. For each complex predicate listed in (1) for English, one example in Catalan will be given, with the exception of strong resultatives and the relevant verb-particle construction, which have already been shown to be absent from Catalan (13).

---

[200] As Snyder (1996: 729) puts it: "The Romance languages have long been noted to contrast with English and other Germanic languages in that they categorically exclude resultative constructions (…). Furthermore, the Romance languages systematically lack direct counterparts to the English verb-particle, make-causative, and double-object dative constructions. Thus, Romance appears to be a strong candidate for a language group in which complex predicates of the English type are systematically excluded."

(14) a. Va   fer   marxar en Joan.   (*make*-cause)

Go-PRES.3SG make-INF leave-INF the Joan

'(S/he) made Joan leave'

b. Vaig   veure   marxar   en Joan.[201]   (perceptual report)

Go-PRES.1SG see-INF leave-INF the Joan

'I saw Joan leave'

c. Va   posar   el llibre sobre la   taula.   (*Put*-locative)

Go-PRES.3SG put-INF the book   on   the table

'(S/he) put the book on the table'

d. Va   enviar   la carta a la Núria.   (*to*-dative)

Go-PRES.3SG send-INF the letter to the Núria

'(S/he) sent the letter to Núria'

e. Li   va   enviar   la carta a la Núria. (double-object dative)

CL-3SG.DAT go-PRES.3SG send-INF the letter to the Núria

'(S/he) sent the letter to Núria'

The existence of the double object construction in Romance languages seems to be dependent on whether the language allows clitic doubling. Catalan permits such a structure: in (14e) the dative clitic (*li*) co-occurs with the lexical indirect object (*to the Núria*) (for discussion on double object constructions in Romance languages, see Uriagereka 1988, Branchadell 1992, Demonte 1994a, b, 1995, and Romero 1997, Rigau 2002, Todolí 2002, a.o.).[202] Like resultatives and verb-particle constructions, one could

---

[201] Note that when *Joan* is replaced by a clitic in both *make*-cause and perceptual report constructions, the clitic can be placed between the two verbal predicates (i), as its English counterpart (*Fred made him leave*, *Fred saw him leave*), and also before the two verbal predicates (ii), unlike English (**Fred him made leave*, **Fred him saw leave*).

(i)   Va fer-lo marxar. (Go-PRES.3SG make-INF+him leave) 'He made him leave'
          Vaig veure'l marxar. (Go-PRES.1SG see-INF+him leave) 'I saw him leave'
(ii)   El va fer marxar. (Him go-PRES.3SG make-INF leave) 'He made him leave'
          El vaig veure marxar. (Him go-PRES.1SG see-INF leave) 'I saw him leave'

[202] While it is usually assumed that Romance languages lack the double object construction, we follow authors like Demonte (1994a: 72) when she claims that "Spanish has the dative alternation or, more strictly, that Spanish sentences with dative clitic doubling share the syntactic and semantic properties of English or German double object sentences". Some examples follow (p. 71-72):

(i)   a. Le   entregué   las llaves al   conserje.
             CL-DAT.3SG give-PST.1SG the keys   to.the janitor
             'I gave the keys to the janitor'
          b. Le   cociné   el pollo   a Mario.
             CL.DAT.3SG cook-PST.1SG the chicken to Mario

also divide double object constructions into those which are relevant to the [TCP], the double object construction present in English, and those which are irrelevant to the parameter, the double object construction present in Romance (which includes a clitic). Even after these concessions have been made, it would not be clear why NN compounding and complex predicates like those exemplified in (14a-d) do apparently exist. If we take the compounding/complex-predicate parameter seriously, languages like Catalan are predicted not to exist. On this view, it is not clear why some complex predicates are available, while others are not if they form a natural class and NN compounding is a possibility in the language.

Catalan is not the only problematic language for Snyder's proposal. It has been reported that Greek has NN compounds and the double object construction, but has neither resultatives nor verb-particle constructions (cf. Horrocks & Stavrou 2007).[203] It seems that complex predicates are not learned as a block, which is in contradiction with Snyder's claim. (Also, as seen above for Basque and Catalan, NN compounding is not a good predictor of complex predicates in Greek).

Another problematic language is Russian, which poses a problem for Snyder's (1995f) implicational relationship from complex predicates to NN compounds. Snyder (1995: 33-34; 2001: 338) shows that Russian allows neither NN compounds like *worm can* nor resultative constructions like *paint the house red*.[204] While the resultative is of the weak type and hence it is irrelevant to the compounding/complex-predicate correlation, Spencer & Zaretskaya (1998) show that Russian does not allow strong resulatives of the English type either.[205] However, they claim that a large class of

---

'I cooked the chicken to Mario'
c. Le   limpié/fregué   las manchas a   la camisa.
   CL.DAT.3SG wipe.off- PST.1SG the stains   from the shirt
   'I wiped the stains off of the shirt'

[203] In Horrocks & Stavrou's (2007: 635) terms: "Greek has the double object construction (Anagnostopoulou, 1994, 2003) and arguably also compounds of the *coffee cup* kind (Ralli and Stavrou, 1997), but it does not have adjectival resultative predication or verb particle complexes (…)".
[204] In Russian a preposition is necessary in the resultative construction as well as in the compound. Consider the following examples taken from Snyder (2001: 338):

(i)   Ivan pokrasil   dom v krasnyj tsvet.
         John paint-PST house in red   colour
         'John painted the house in the color red'
(ii)   banka dlja chervej
         can   for worms

[205] The Russian counterpart of the strong resultative in the English sentence *The river froze into a block of ice* is ungrammatical.

Russian prefixed verbs possesses the same semantic characteristics as the English strong resultatives (e.g. *They drank the pub dry*).[206] They argue that the prefix in Russian has the same function as the adjective in the resultative construction in English (see also Acedo-Matellán 2009). If this is correct, then Russian is a real threat to the implicational relationship from resultatives to NN compounds.

Latin causes a problem similar to that caused by Russian. Latin lacks productive NN compounding and adjectival resultatives but unexpectedly for Snyder (1995f), it has the double object construction and the resultative construction expressed by means of a prefix (cf. Acedo-Matellán 2009).

Given that Snyder leaves room for the possibility that in [+TCP] languages not all complex predicates must necessarily be available, Catalan and Greek could be argued to fit into this possibility. This strategy, though, really weakens Snyder's proposal and still cannot account for languages like Russian and Latin, which have complex predicates like strong resultatives (by means of a prefix) and no productive NN compounding.

Son (2007) also presents a number of languages as counterexamples to Beck & Snyder's (2001a) claim that there is a strong correlation between strong adjectival resultatives (e.g. *hammer the metal flat*) and telic path PPs (*run to the store*), the latter being interpreted as a result phrase. Beck & Snyder argue that both constructions should pattern together: the availability of adjectival resultatives predicts the availability of telic path PPs and vice versa.[207] According to Son (2007), there is no necessary correlation between the availability of both constructions in some of the languages studied by Snyder (2001) (Korean, Japanese and Hebrew) as well as in other languages

---

 (i) *Reka zamërzla v glybu l'da.
     River froze into block of-ice

[206] Consider the example in (i) which is given the lexical conceptual structure (LCS) in (ii) (cf. Jackendoff 1990), which is parallel to the LCS given for *They drank the pub dry* in (iii). These examples are drawn from Spencer & Zaretskaya (1998: 17, ex. (51) and (52); 7, ex. (19)).

 (i) Ona is-pisala svoju ručku.
     She IZ-write her pen-ACC
     'Her pen has run out the ink'
 (ii) [[CAUSE[ACT (*she*)], IZ (*pen*)], BY [WRITE (*she*)]]
 (iii) [[CAUSE[ACT (*they*)], BECOME [DRY (*pub*)]], BY [DRINK (*they*)]]

[207] Beck & Snyder's (2001a: 118) conclusion to their study is as follows: "(…) languages that clearly disallowed the resultative never permitted a goal-phrase to convert an activity into an accomplishment, and languages that clearly permitted the resultative always allowed the goal-PP construction to be interpreted as an accomplishment".

(Indonesian and Czech). Son's examples, though, have to be carefully examined before one can conclude that they are real counterexamples to Beck & Snyder's proposal.

For example, regarding Hebrew, it is denied that there are adjectival resultatives and goal PP constructions in Beck & Snyder. Son (2007: 139-141, ex. (25), (27), (28), (29))[208] provides a few examples showing that goal PPs are indeed grammatical in Hebrew. While these examples may be grammatical, they do not show that an accomplishment has been created (e.g. the in-x-time adverbial test is not applied). Interestingly Horrocks & Stavrou (2007: 609) argue that goal PPs in Hebrew are adjuncts and as such they do not change the aspectual character of the verb. On their view, unaccusativization of verbs of manner of motion does not take place in Hebrew. As they put it (p. 609), "(…) the verb retains its basic activity/manner of movement sense, as the rejection of time-within-which adverbials makes clear, and the goal PP serves only as an adjunct, specifying the arbitrary end-point of the particular movement involved (…)". The validity of their claim can be observed from the impossibility of an in-time adverbial co-occurring with a goal-phrase in Hebrew.[209] Whether strong adjectival resultatives are a feature of Hebrew is not obvious from the examples provided by Son (2007: 138, ex. (22))[210] because they contain weak resultatives. Thus, the correlation between adjectival resultatives and path PPs cannot be questioned, contra Son (2007).

In the case of other languages, we agree with Son's conclusion that resultatives cannot be treated as a uniform group. She observes that Japanese speakers disagree on the acceptability of resultatives, namely *John wiped the table clean* is accepted while *Mary beat the metal flat* is not. Such disagreement is expected if, as discussed above, Japanese has no strong resultatives (cf. Washio 1997, Tomioka 2004), and one

---

[208] Consider the following example from Son (2007: 140, ex. 25a).

(i)   David {rac/zaxal}   el-tox   ha-xeder.
        David   ran/crawled ALL-inside the-room
        'David ran/crawled into the room'

[209] Horrocks & Stavrou (2007: 609) provide the following example (from Beck & Snyder 2001b).

(i)   *Dan halax   el ha-kfar   tox Sa'a.
         Dan walked to the-village in   an-hour

[210] Consider the following weak resultative provided by Son (2007: 138, ex. 22b)

(i)   *Hu cava   et   ha-kir   adom.
         He painted ACC the-wall red
         'He painted the wall red'

resultative to be tested is of the strong type (*Mary beat the metal flat*). The fact that Japanese has only weak resultatives is inconsequential for the compounding/complex predicate correlation since this is only sensitive to strong resultatives but predicts that goal PPs should not be available. Snyder (2005) argues that despite the limited number of adpositions, which explains why Japanese lacks verb-particle constructions, the language nevertheless does allow constructions with goal PPs.[211] This conclusion cannot be confirmed since the only example he provides is not used with the in-x-time adverbial, although it can be contrasted with Levin & Rappaport Hovav's (1995: 185) position, according to which Japanese can only take goal PPs if the verb (e.g. *run*, *swim*) combines with *go*. On their view, Japanese allows telic PPs only if complex verb formation including an inherently directional verb is present.[212] Unfortunately, the examples provided by Levin & Rappaport Hovav do not come with an in-x-time adverbial. Given these contradicting claims and in the absence of relevant data, the purported correlation between strong resultatives and goal PPs cannot really be questioned. If complex predicates were learned as a block, as Snyder claims, the fact that strong resultatives do not depend on adpositions and yet are absent from the language could suggest that complex predicates may be generally missing from the language (goal PPs included). If Japanese had neither strong resultatives nor goal PPs, then the aforementioned correlation could be said to hold (contra Son). We leave the question of whether Japanese has real goal PPs (in the sense that they unaccusativize the verb) to future research.

In short, some of Son's alleged counterexamples to Beck & Snyder's (2001a) correlation between adjectival resultatives and goal PPs may not be real counterexamples once the data are examined more carefully. They can, nonetheless, be used to re-examine the values given (by Snyder 2001) to the adjectival resultatives in

---

[211] Snyder (2005: ex. 56) provides only one example with a goal PP in Japanese, which is given below.

(i)   hasi-no   sita-e   oyogu-no-wa zikan-ga kakarisugitu.
        Bridge-GEN underneath-TO swim-ING-TOP time-NOM take.too.much
        'Swimming under the bridge takes too long'

[212] Consider the contrast between (a) and (b) below (examples from Yoneyama 1986: 1-2, ex. 1b, 3b).

(i)   a. ?John-wa kishi-e   oyoida.
          John-TOP shore-to swam
        b. John-wa kishi-e   oyoide-itta.
          John-TOP shore-to swimming-went
         'John swam to the shore'

Table 3.1 (section 3.1). For example, Japanese has been shown not to have the resultative relevant to the [TCP], with the result that the value given in Table 3.1 should be revised (as already discussed in subsection 3.1.3).[213]

From the discussion above we can conclude that, when more and more languages are studied in depth, the apparently exceptional behaviour of Basque in being a [+TCP] language and yet not having strong adjectival resultatives is not so exceptional. Languages like Catalan, Greek and Japanese also challenge a strict application of the compounding/complex-predicate parameter in being [+TCP] languages and not allowing the resultative construction. Although this is a problem for a strict application of Snyder's parameter, he could still argue that in those cases the positive value of the TCP is not enough and that there must be another factor (or other factors) explaining the non-existence of resultatives (which is what he actually argues for Basque). Crucially Snyder's (1995f) alleged dependence of complex predicates on NN compounding is also challenged: Latin and Slavic languages like Russian have resultatives (expressed by means of a prefix) but no NN compounding. Additionally, Snyder's claim that complex predicate constructions are learned as a block has to be discarded in the view that languages like Catalan, Greek, Russian and Latin allow only some of them. This last point further questions the compounding/complex-predicate parameter, whose effect seems to be nullified after all the counterexamples just mentioned.

In short, the main conclusion to be drawn from this section is that for Snyder's parameter to work one must first identify which complex predicates are relevant to the [TCP]. As we have seen, the [TCP] must be sensitive only to strong resultatives (vs. non-strong resultatives), non-directional verbs in verb-particle constructions (vs. directional verbs), double-object constructions without clitic doubling (vs. those with clitic doubling), etc. In other words, the group of complex predicates is not given a precise characterization. Even after this clarification has been made, though, the [TCP] is not without problems. We have challenged the link between adjectival resultatives and NN compounding, the implicational relationship of complex predicates to compounding and the interdependence among complex predicates.

---

[213] The resultative available in Korean also seems to be different from the English-type resultative (Shim & den Dikken 2007: 21), with the consequence that the value given by Snyder (2001) to the resultative in Korean should also be revised in Table 3.1 (section 3.1).

Assuming that there is a link between compounding and resultatives - as Snyder does, although not in the way he proposes - in the next section two syntactic accounts of the cross-linguistic variation found in resultatives will be briefly reviewed. We will consider whether these two accounts can be made extensible to compounding and whether they can explain why in some languages like Catalan NN compounds are productive although to a lesser degree than NN compounds in English, a question which will be specifically addressed in subsection 3.2.2.

3.2 Breaking down the Compounding Parameter

While the validity of the [TCP] and its implications have been questioned in the previous section, let us assume now that the alleged link between resultatives and compounding is real, as Snyder (1995f) does. In order to identify the connection between the two constructions, two alternative analyses of resultatives will be presented.

3.2.1 Some remarks on resultatives

In the literature, there are quite a few syntactic analyses of resultative constructions available. Even though most of them are not designed to explain compounding (e.g. Hoekstra 1988, Neeleman & van de Koot 2002b), some attempts are made to link the availability of resultatives to the availability of compounding. Two recent accounts will be briefly reviewed here: Kratzer's (2005) and Mateu's (2000, 2010). The details of each analysis will not be presented. For our purposes it will suffice to summarize the core ideas of the two accounts in the following subsections.

3.2.1.1 Kratzer (2005)

Kratzer presents a syntactic analysis of the adjectival resultative, an example of which is shown in (15), which allows her to explain why such constructions (and by extension primary compounds) are available in some languages while impossible in other languages. In her proposal the morphological properties of the adjectival element play a crucial role since they determine whether adjectival resultatives (and primary compounds) are an option in a language.

(15)   die Teekanne leer   trinken   (German)

       the teapot   empty drink

       'To drink the teapot empty'

Kratzer (2005: 177, ex. 1)

Kratzer follows Hoekstra (1988) in adopting a raising analysis for adjectival resultatives, which is depicted schematically below (taken from Kratzer 2005: 180, 203).

(16)

```
                        VP
                       /  \
                     VP     V
                    /  \   trinken
                  AP     V
                 /  \  [cause]
               DP     A
          die Teekanne  leer
```
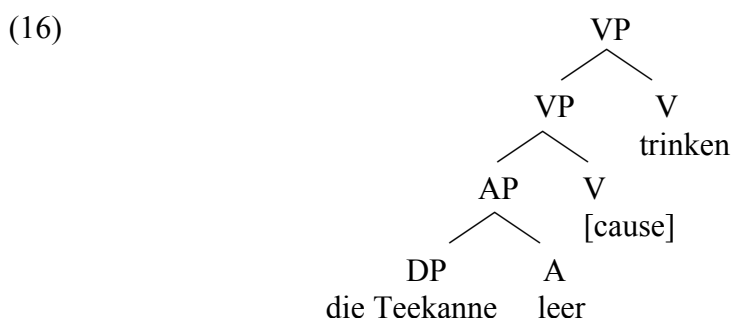
In order to understand (16), some comments are necessary. *Die Teekanne* is an argument of the adjective *leer* (not of *trinken*). Kratzer assumes that "the unique (non-eventuality) arguments of the adjectives are internal" (p. 179) and that in the syntax arguments originate within the projection of their heads, which explains why *die Teekanne* is within the projection of *leer* in (16). Although *die Teekanne* is placed in its base position (within the projection of its head, *leer*), it will move into the functional structure of the verb *trinken* for case reasons and thus become the surface direct object of the compound *leer trinken*. Our main focus, though, will not be on the raising of *die Teekanne* but on the structure between *leer* and *trinken*, both elements included.

Kratzer assumes that an unpronounced affix with the interpretable feature [cause] is responsible for the causal relation present in adjectival resultatives (16). Kratzer treats [cause] as a derivational affix and assumes, following Pylkkänen (2002), that [cause] introduces an event argument. Kratzer further assumes that [cause] is also responsible for turning adjectives into verbs, i.e. for the change of stative roots into eventive predicates in deadjectival verbs like those in (17).[214] The affix may be unpronounced (17a) or pronounced as *–en* (17b).

---

[214] For a similar view, see Padrosa-Trias (2005b, in press, a), where it is argued that a suffix, which can be full (-*en* suffix) or empty (zero-suffix) of phonological content, is responsible for the change of category.

(17)   a. empty, dry, clean, cool, dim, dirty…

        b. flatten, shorten, blacken, sweeten, stiffen…

Kratzer (2005: 201, ex. 65)

From what has been said so far, the contrast in grammaticality of the following examples is unexpected.

(18)   a. The gardener watered the tulips flat.

        b. *The gardener watered the tulips flatten.

Kratzer (2005: 201, ex. 66)

Kratzer resolves the contrast between the adjectival resultative in (18a) and the verbal causative in (18b) by means of the distinct composition of the suffix involved in the two constructions. It is claimed that the unpronounced suffix present in *flat* (18a) spells out only [cause] while the suffix *–en* visible in *flatten* (18b) is assumed to spell out two features: [cause] and a voice feature which can be [active] or [non-active] depending on whether it derives the transitive or intransitive version. The ungrammaticality of (18b) follows from the assumption that voice features are inflectional and that the compound formed by an adjectival root and the [cause] affix cannot put on any inflectional morphemes in the resultative. This constraint is based on Baker's (1996, 2003) 'Proper Head Movement Generalization', which is given the following definition.

(19)   "The Proper Head Movement Generalization

        It is impossible to move from a functional category into a lexical category."

Baker (2003: 306)

In accordance with (19), head movement can proceed as it should in (18a) but not in (18b) where head movement should stop when it hits the feature [non-active], which as a functional head should prohibit movement into the higher lexical category WATER. The resulting sequence of head adjunction needed for (18b) would look like as in (20).

(20)   FLAT+[cause]+[non-active]+WATER

The head movement chain in (20) should be read as follows. First the lexical head FLAT moves to another lexical head, the silent derivational affix [cause]. In the second step, the complex head FLAT+[cause] adjoins to the functional head [non-active]. The illicit movement comes when in the next step the complex head just created moves to a lexical category (WATER) since it moves from a functional category. By contrast, in the case of simple stative roots like *flat* (18a), the feature [non-active] is absent and as a result the complex lexical head FLAT+[cause] can move to the higher lexical head, WATER, without incurring any violation of the constraint built on the generalization in (19).

A similar account could explain why adjectives like *disgusting* and *shined* cannot appear in adjectival resultatives (21), assuming that *–ing* and *–ed* are inflectional morphemes. Such morphemes will intervene between two lexical heads. For instance, in the case of *disgusting*, *-ing* will come between DISGUST and [cause]. The illegal movement will occur when the complex head created by attaching DISGUST to *–ing* has to move from a functional head to the lexical head [cause].[215]

(21)   a. *I cooked it disgusting.

    b. *I brewed it soothing.

    c. *The maid scrubbed the pot shined/shining.

    d. *The tourists walked their feet blistered.

Kratzer (2005: 200, ex. 63, 64)

On Kratzer's account the availability of adjectival resultatives in a language crucially depends on the absence of inflectional material on the adjectival root. This explains why adjectival resultatives are available in English and German while absent in Romance languages since in languages like French adjectival roots cannot appear by themselves. Agreement morphology (e.g. gender, number) is necessary on adjectives independently of whether they are used attributively or predicatively. If adjectives in Romance are inflected from the beginning of the syntactic derivation, moving from a functional category to a lexical category would be unavoidable in the formation of adjectival

---

[215] If *–ing* and *–ed* happened to be derivational, then the ungrammaticality of the sentences in (21) could be attributed to ordering constraints between affixes. For example, according to Hay (2000, 2002: 528), "an affix that can be easily parsed out should not occur inside an affix that cannot." Given this account, it would seem plausible that *–ing* and *–ed*, which can be easily parsed out, occurred outside the unpronounced affix [cause], assuming that unpronounced elements are difficult to parse.

resultatives, thus violating the constraint on head movement mentioned above. This is Kratzer's account to explain why adjectival resultatives are absent in Romance.

Kratzer links her account of adjectival resultatives to Snyder's (1995f) correlation between the availability of resultatives and 'root' compounding (primary compounds in our terms).[216] On Kratzer's account, the link between adjectival resultatives and compounding is based on the fact that serial verb constructions and VV compounds are structurally similar (with the possibility of a serial verb construction giving rise to a VV compound or involving covert compounding) and that adjectival resultatives are "the closest analogue to a serial verb construction" in languages like English and German (p. 203). For example, *trinken* and *leer*+[cause] in (16) are two independent eventive predicates which are placed next to each other without a visible coordinating or subordinating marker, properties characteristic of serial verb constructions. From these facts, Kratzer concludes that "whatever forces compounding for serial verb constructions can be assumed to force compounding for adjectival resultatives as well" (p. 204). From this statement, it seems that if inflectional morphemes are not permitted in adjectival resultatives, they should also be prohibited from appearing in compounding.

What type of compounding does Kratzer have in mind, though? Compounds with an adjective as one of their constituents seems to be the most probable answer because it would establish a connection between the adjective in adjectival resultatives and the adjective in compounding. 'Bare roots' are predicted to be the only elements allowed in adjectival resultatives and in compounds with one adjectival constituent. Let us consider how this prediction fares in English. That adjectival resultatives in English only permit bare adjectival roots is clear from the data above (18, 21) (but see the comment on roots vs. lexemes below). As for compounding, chapter 2 contains compound types like the [AN]$_N$ type (e.g. *red squirrel* (38a)) and the [AV]$_V$ type (e.g. *deep-fry* (51)) which would satisfy the requirement if a root is understood as being a base without inflectional morphology, which seems to be Kratzer's understanding of the term. (Recall from chapter 2 that in our understanding, words like *red* and *deep* are not roots but lexemes, bases ready to enter syntax.) These two compound types do not

---

[216] Recall from chapter 2 (section 2.2.2) that root compounds are generally understood as those compounds in which the non-head is not an argument of the (de)verbal head or the head is not (de)verbal. Examples of root compounds include forms like *mass production, fitness campaigner, climbing equipment, amusement park* and *retirement age*. Since compounding elements can be complex words including suffixes like *–er*, *-ness*, *–ment* and *–tion*, we find the term 'root compounds' inappropriate, which we replace with the expression 'primary compounds'.

question Kratzer's prediction according to which inflectional morphology cannot appear on compounding adjectives. However, on Kratzer's account there is no explanation for languages whose adjectives can appear in a bare form when they are used as the first constituent in a compound and yet do not allow adjectival resultatives of the English type (e.g. *beat X flat*). This is the case for Greek, a highly inflected language, in which adjectives usually show agreement with the nouns they modify. In Greek, bare adjectival forms can appear as the first constituent in a compound but cannot form part of an adjectival resultative (on this point, see Horrocks & Stavrou 2007: 636).[217] In addition, Kratzer's analysis cannot explain why adjectival resultatives are possible in Icelandic since the adjective bears agreement morphology (see Whelpton 2007 for examples and discussion and Acedo-Matellán 2009 for an alternative analysis of resultatives which captures the facts reported for Icelandic).[218]

Let us now consider another possibility. Since Kratzer intends to relate the absence of inflectional morphology on the compound formed by an adjectival resultative to Snyder's (1995f) correlation between the availability of resultatives and 'root' compounding (primary compounds in our terms), NN primary compounds could be the relevant compound type for Kratzer. If this were the case, NN compounding should be submitted to the same restrictions as adjectival resultatives: inflectional morphology should be absent from such compounds. By looking at NN compounding in English (see for example the compounds of (32c) and (35) in chapter 2), most of them

---

[217] Horrocks & Stavrou (2007: 636) observe that "adjectives of the type *neos* 'new' (nominative masculine singular) have a feminine *nea* and a neuter *neo* (<earlier *neon*), each of which has distinct forms for the nominative, accusative and genitive cases in both the singular and the plural. But when such adjectives appear as the first element of compounds, they have no inflectional suffixes at all, appearing as *neo-*, etc. in both ancient and modern Greek. Note that this element is therefore distinct in ancient Greek from the corresponding neuter singular, and that the homophony in modern Greek is simply the accidental product of the quite general loss of final /-n/." They provide the following examples:

(i)   a. neo-plutos   'nouveau-riche'   (modern Greek and ancient Greek)
       b. thermo-emos   'hot-blooded' (temperamental)   (modern Greek)

However, Horrocks & Stavrou also note that bare adjectival forms cannot be part of resultatives like *epipedho-xtipo* 'flat-beat'.

[218] Whelpton (2007: 3, ex. 7, 13) provides examples like (i), which clearly shows that the adjective is inflected. This contrasts with overt compounding, where adjectives appear uninflected, as in (ii).

(i)   Járnsmiðurinn   hamraði   málminn   flatan
       blacksmith.the hammered metal.the-MASC.ACC.SG   flat-MASC.ACC.SG
       'The blacksmith hammered the metal flat'
(ii)   hreinskrúbbuðu   pönnurnar
       clean-scrubbed-FEM.NOM.PL   pans-FEM.NOM.PL.the-FEM.NOM.PL

seem to conform to the restriction just mentioned although some compounds might question it. Consider some of the compounds included in (32c) of chapter 2.

(22)   a. advice centre, apron string, bar code, cane sugar, chicken fat, and picture book.

b. chewing gum, cleaning lady, drinking water, ironing board, lodging house, reading glasses, and reading material.

c. fees controversy, suggestions box, Universities yearbook, and weapons system.

Compounds like those in (22a), which are very common in English, do not pose any challenge to Kratzer's account: there is no inflection on them. As for the compounds in (22b), the nature of the suffix *–ing* is not obvious. It is not clear whether *–ing* should be equated with the derivational suffixes present in compounds like *amusement park*, *correspondence course* and *detention centre* or should rather be treated as an inflectional suffix (our own inclination). If the second option proves true, then Kratzer's analysis is questioned. Regarding the compounds in (22c), the status of the suffix *–s* is also controversial in the literature on compounding. According to authors like Olsen (2000b: 898), the *-s* should not be taken as a plural inflection,[219] but in other authors' opinion, *-s* clearly marks plural. For example, Pullum & Scholz (2002: 24-26) provide a few examples of compounds of which they argue that the non-head position bears plural inflection. Consider the following compounds: *chemicals-maker*, *forms-reader*, *generics-maker*, *securities-dealer*, *drinks trolley*, *rules committee*, *publications catalogue*, *parks commissioner*, *programs coordinator*, *buildings inspector*, *faces research*, *letters policy*, *complaints department*, *claims backlog*, *counterexamples list*, etc. To make the point clear, Pullum & Scholz (p. 25) observe that "it is perfectly possible to have a factory making just one chemical, but a chemicals factory makes more than one, just as a forms reader reads arbitrary different forms rather than just one form, and so on." In short, the examples just considered provide strong evidence for the claim that *–s* instantiates plural inflection. If inflectional morphology is present in the case of NN compounds, Kratzer's alleged relation between adjectival resultatives and

---

[219] Olsen (2000b: 898) states that "they [determinative compounds] do not allow inflectional morphemes marking grammatical categories of case, number, person etc. internally (*punchcard* vs. **punchedcard* (…)."

NN compounding cannot hold. In her analysis, adjectival resultatives crucially do not allow inflectional morphemes.

Even if NN compounds were the relevant type of compound in Kratzer's account, we find no connection between the relevance of adjectives being uninflected in adjectival resultatives and the availability of NN compounding, since the latter do not involve adjectives. That is, it is not clear to us why uninflected adjectives in resultatives should matter for the formation of NN compounds in a language. In short, the connection between the two constructions does not seem to be well-founded. In addition, the syntactic analysis proposed for adjectival resultatives cannot be applied to compounds if we are right in deriving compounds in a morphological component, which is different from the syntactic component (see section 1.2 in chapter 2 for some evidence of our position).

Leaving aside the alleged connection between resultatives and compounding, Kratzer's account of adjectival resultatives cannot explain why some non-strong adjectival resultatives are permitted in Catalan since agreement morphology is present in the adjective involved in this construction (see footnote 186). That is, Kratzer's account predicts that all types of resultatives, independently of whether they are weak or strong (cf. Washio 1997, Kaufmann & Wunderlich 1998), should be non-existent in languages where adjectives display inflectional morphology. As noted, this prediction is falsified by languages like Catalan (and other Romance languages) which allow non-strong resultatives despite the adjective being inflected. Crucially, the problem of the adjective being inflected was also found in strong adjectival resultatives in Icelandic.

In conclusion, Kratzer's account of adjectival resultatives in English and German does not seem to be applicable cross-linguistically, as it is intended. In addition, Kratzer's attempt to link her account of adjectival resultatives to Snyder's (1995f) correlation between adjectival resultatives and 'root' compounding fails because she seems to understand the term 'root' differently from Snyder. For Kratzer, a root is a base without inflectional morphology with the apparent consequence that root compounding should not have inflectional material, while Snyder understands root compounds as they are usually interpreted, namely as primary compounds (see footnote 216 and the examples in (22)).

3.2.1.2 Mateu (2000, 2010)

Within the framework of H&K's (1993, 1998) syntactic theory of argument structure,[220] Mateu (2000) presents an analysis of resultatives to explain why they are available in English (23) but not in Romance (24). He explicitly says that the parametric variation involved in the resultative construction applies only to strong resultatives (cf. Washio 1997, Kaufmann & Wunderlich 1998). Non-strong resultatives, which are available both in English and in Romance, are claimed to be adverbial modifiers. As such, they are placed outside the main argument structure of the sentence, i.e. they are VP-adjoined. By contrast, strong resultatives are argued to be internal to the main argument structure of the sentence (see below).

(23)   a. John hammered the metal flat.
        b. The dog barked the chickens awake.

                                            Mateu (2000: 72-73, ex. 1-2)

(24)   a. *En Joan martellejà el metall pla.   (Catalan)
        b. *El gos bordà els pollastres desperts.

                                            Mateu (2000: 73, ex. 4)

Mateu derives the resultatives[221] in (23) by adopting the lexical subordination approach, initiated by Levin & Rapoport (1988) (see also Jackendoff 1990, 1992, Spencer & Zaretskaya 1998), in which the result phrase is taken as the core predicate (*flat* in (23a)) and the verb as the subordinate predicate (*hammer* in (23a)). According to Mateu, Talmy's (1985, 1991) work on lexicalization patterns provides the evidence for a syntactic approach to the lexical subordination process present in resultatives. In this respect, Talmy's distinction between satellite-framed languages and verb-framed languages is crucial since the lexical subordination process, and by extension resultative formation, is possible only in the former. Let us see why. A language is classified as satellite-framed or verb-framed depending on which semantic components (e.g. 'figure', 'path', 'manner', 'cause', 'motion', etc.) are conflated into the verb. Conflation of motion with manner is typical of satellite-framed languages like English, German and Dutch, where the path relation remains as a satellite around the verb (25). By contrast,

---

[220] H&K's (1993, 1998) theory of Lexical Syntax will not be reviewed here. Only will some points essential to the understanding of Mateu's analysis be presented.
[221] As earlier, 'strong resultatives' are referred to as 'resultatives' when the two types of resultatives need not be differentiated.

verb-framed languages like Catalan, Spanish, French and Japanese typically have conflation of motion with path. Such conflation has a fossilized nature, thus saturating the verb lexically and preventing the manner component from being conflated into the verb.[222] The manner component can only be expressed as an adjunct (26).

(25)   The boy danced into the room.

(26)   El   noi va entrar   a   l'habitació ballant   (Catalan)
        The boy went+into loc-prep the room   dancing

                                                      Mateu (2000: 85, ex. 21)
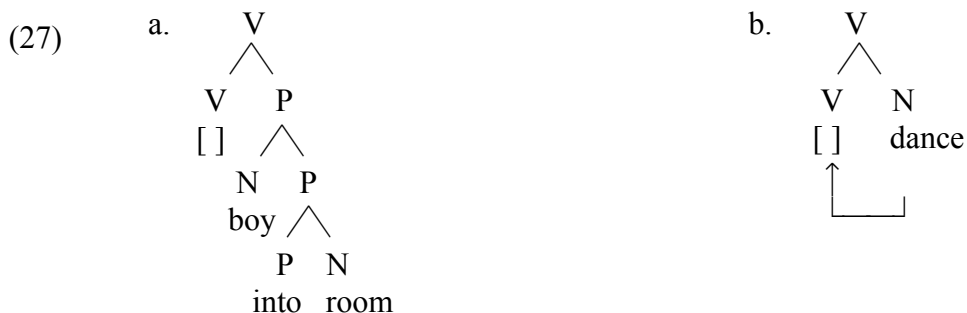
Mateu proposes that the properties associated with the element expressing directionality/path are responsible for the parameterization of the conflation processes observed in (25)-(26): the satellite nature of the element encoding the path relation in English allows the language to conflate motion with manner (25), whereas the fossilized conflation of motion with path in Catalan prohibits the language from conflating motion with manner (26). Mateu provides these conflation processes with a l(exical)-syntactic analysis, of which a sketch follows. The sentence in (25) is derived by conflating the two independent l-syntactic structures depicted in (27): (b) is subordinated to (a). The l-syntactic structure of (27b)[223] is able to conflate into the verb in (27a) because the verbal matrix is phonologically empty (and phonologically empty matrices must be avoided at PF) and because the element expressing path (*into*) is not lexically encoded into the verb.[224]

---

[222] The fossilized nature of this conflation process can account for the lack of elasticity of the verb meaning in Romance, as opposed to languages like English where such elasticity is possible (Rappaport Hovav & Levin 1998).
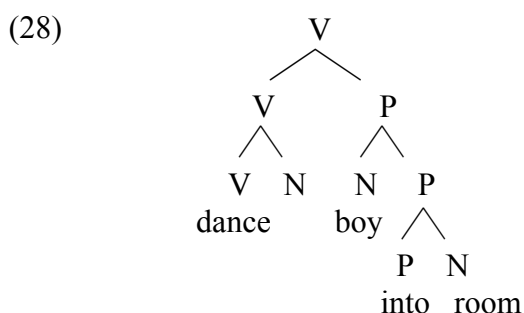
[223] Unergative verbs like *dance* are assumed to be denominal verbs. For a criticism of this view, see Don (1993).

[224] The first part of this subsection is based on Mateu (2000), but bear in mind that a similar analysis can be found in Mateu (2001) and Mateu & Rigau (2002) with minor modifications. In Mateu & Rigau (2002), for example, the verbal matrix of (27a) is not completely empty but contains an abstract predicate, 'go', to which the phonological matrix of the unergative verb *dance* is adjoined by means of Merge. Since the path component is lexically incorporated into the verb in Romance, there is no empty predicate 'go' to which a manner verb like *dance* can be adjoined. The fact that a prepositional complement can appear with verbs which lexically encode directionality (Cat. *entrar* + PP 'to enter') is explained by the claim that "P is always projected in syntax, this being a copy of the P incorporated into the verb" (p. 224). For English sentences like *John danced Sue into the room* another abstract predicate, 'cause', is postulated. The same abstract causative predicate is argued to be involved in sentences like (i)-(vi) below (Mateu & Rigau 2002: 221).

    (i)   Sue danced the night away.

(27) a.

```
          V
         / \
        V   P
       [ ]  / \
           N   P
          boy  / \
              P   N
            into room
```

b.

```
          V
         / \
        V   N
       [ ]  dance
        └────┘
```

After the relevant conflation processes have taken place, the resulting l-syntactic structure is as follows.

(28)

```
              V
            /   \
           V     P
          / \   / \
         V   N N   P
       dance boy  / \
                 P   N
               into room
```

Mateu (2000: 86, ex. 22, 23)

The same analysis is extended to adjectival resultatives. Let us consider the plausibility of such an extension. Mateu treats the P, as in (28), as a cover birelational term for adpositions, adjectives and adverbs. Adjectives are treated as a composite unit which is the result of conflating a non-relational element into a relational one (cf. chapter 1, section 1.1; see also Mateu & Rigau 2002, Amritavalli & Jayaseelan 2003, Kayne 2009). In addition, Ps and adjectives are assigned the same conceptual structure: Ps encode physical space and adjectives are associated with abstract space (Jackendoff 1990). Also, that Ps and adjectives may be instances of the same object nicely fits in with the fact that English has both adjectival and prepositional resultative constructions while Romance languages lack both of them.[225] This fact can be captured if adjectival

---

    (ii)   Tribal members ceremonially danced it open.
    (iii)   Sue laughed herself silly.
    (iv)   Sue sneezed the napkin off the table.
    (v)   Sue laughed her way into the room.
    (vi)   Sue swam her swimsuit to tatters.

[225] Mateu (2000: 79) presents the following set of data. The examples in (iii) and (iv) are the exact Catalan counterparts of the English adjectival and prepositional resultatives in (i) and (ii) respectively. The Catalan sentences are grammatical in their non-resultative readings (i.e. when the adjective *open* has

and prepositional resultatives involve the same element encoding path, covert in the first case and overtly expressed by means of a preposition in the second case. In conclusion, adjectives seem to involve an abstract path relation, with the consequence that adjectival resultatives can plausibly be given the same analysis as that depicted in (28). In l-syntactic terms, the adjectival resultative in (23a), repeated below, would result from the conflation of the two structures in (29).

(23)   a. John hammered the metal flat.

(29)   a.
```
            V
           / \
          V   X
          []  / \
             N   X
            metal / \
                 X   Y
                 []  flat
                 |___|
                  ↑__|
```
        b.
```
            V
           / \
          V   N
          []  hammer
          ↑___|
          |___|
```

Because the abstract path (*flat*) can be left stranded in satellite-framed languages, an independent structure (29b) can be conflated into the phonologically empty verbal matrix of (29a). The resulting l-syntactic structure is shown in (30).

(30)
```
                    V
                   / \
                  /   \
                 V     X
                / \   / \
               V   N N   X
            hammer metal / \
                        X   Y
                        flat
```

---

an attributive reading and the PP *inside the bathroom* has a locative reading), irrelevant readings for our present concerns.

(i)   Joe kicked the door open.
(ii)   Joe kicked the dog into the bathroom.

(iii)   *El Joe colpejà la porta oberta.
(iv)   *El Joe colpejà el gos a dins el bany. (a dins = inside)

See also the contrast between the examples in (23) and (24).

On this account the absence of prepositional and adjectival resultatives in Romance is explained by the fact that the path relation is always conflated into the verb in these languages, which prevents an independent l-syntactic structure from being conflated into the verb. On Mateu's account, linguistic variation is dealt with by an l-syntactic version of the so-called lexical subordination process, whose application is subject to the non-conflated nature of the element encoding path.

Mateu's analysis of resultatives, unlike Kratzer's, takes into account Washio's (1997) distinction between strong and non-strong resultatives (see also Kaufmann & Wunderlich 1998). His account is designed to explain strong resultatives, the type of resultative which is subject to parametric variation (present in English but absent in Romance) (see footnotes 186 and 187).

The l-syntactic account of resultatives presented so far, though, is not intended to be applied to primary compounds. If a link could be established between resultatives and primary compounds, it would have to be between weak resultatives and compounding since strong resultatives are absent from Romance and primary compounds have been shown to be available in Catalan. In addition, one would expect the availability of the two objects (weak resultatives and compounding) to hinge on the same fact: the syntactic version of the lexical subordination process presented above for strong resultatives should not be able to take place. By contrast, it seems that an element encoding path would have to be assumed, which would be in line with Mateu's assumption that the parameterization of Talmy's conflation processes have to do with the properties associated with the element expressing directionality/path. Path has been shown to have a fossilized nature when conflated with motion in Romance. At this point the question of what the element encoding path would be in compounding arises. It is not obvious what the answer to this question would be since no path element seems to be involved in most compounds (e.g. *barca cisterna* (boat tanker) 'tanker boat', *home aranya* (man spider) 'a spiderman'). In short, it seems that no link can be established between weak resultatives and compounding. Mateu's (2000) account of resultatives, though, cannot be blamed for not providing the link since it is not designed to explain compounding. A connection seems to be made in this author's more recent work, though (Mateu 2008, 2010, Mateu & Rigau 2009, 2010).
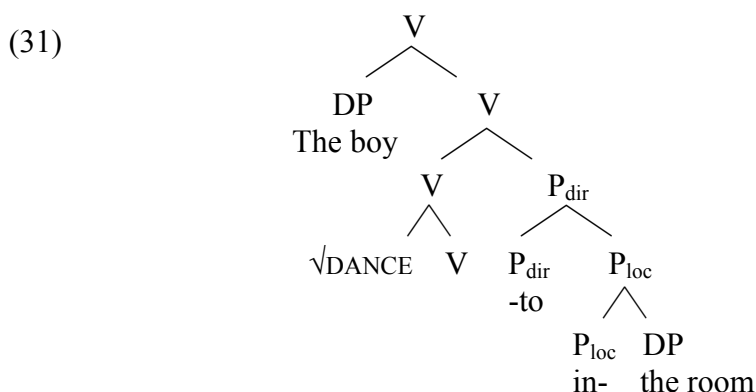
Mateu (2010) and Mateu & Rigau (2010) treat strong resultatives as being instances of conflation/compounding whereas weak resultatives are treated as instances

of incorporation. Let us now consider how these authors reach this conclusion and whether their analysis can be extended to primary compounds.

Mateu (2008, 2010) and Mateu & Rigau (2009, 2010) present a slightly different analysis for sentences like those in (25) and (26), repeated below. The new analysis for (25) is depicted in (31). As can be seen, a root (√DANCE), instead of an unergative verb, is subordinated to a null transitive verb. The complex verb is now created by means of a root-verb compounding (cf. McIntyre 2004, Zubizarreta & Oh 2007). Again, conflation of motion (the phonologically empty verb) with manner (the root √DANCE) is possible due to the non-conflating nature of *into*.

(25)   The boy danced into the room.

(26)   El   noi va entrar   a   l'habitació ballant   (Catalan)
        The boy went+into loc-prep the room   dancing

(31)

```
                          V
                        /   \
                     DP        V
                  The boy    /   \
                           V       P_dir
                          / \     /    \
                  √DANCE   V  P_dir   P_loc
                             -to      /  \
                                   P_loc  DP
                                    in-  the room
```

By contrast, the sentence in (26) has a different derivation (32). In Romance the manner component can only be expressed by means of an adjunct (not represented in (32)) because the path element has been conflated into the verb (conflation of motion with path in Talmy's 1985, 1991 terms), thus giving the verb a directional meaning and phonological content (*entrar* 'to go in'). In verb-framed languages then, no independent root can be merged with the phonologically abstract motion verb (i.e. the root-verb compounding strategy is not available). The direction encoded into the verb can be

further specified by an additional directional P, like *a l'habitació* (26/32).[226] It is assumed that the visible directional P is inserted into P_{loc} after the verb has been formed (i.e. after an abstract P has been conflated into the verb).

(32)

```
                    V
                  /   \
               DP       V
              El noi   /   \
                      V        P_dir
                    /   \      /    \
                 P_dir   V   P_dir   P_loc
                      va entrar     /    \
                                 P_loc    DP
                                  a    l'habitació
```
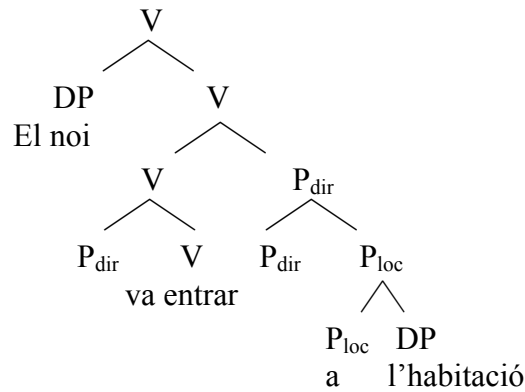
Details aside, the analyses presented in (31) and (32) are applied to strong and weak PP/AP resultatives respectively in Mateu (2010). Adjectives are assumed to involve an abstract path (see the comments above). Mateu adopts and applies to resultatives Haugen's (2009) distinction between conflation/compounding, which is carried out by means of Merge, and incorporation, understood as head-movement (Baker 1988, H&K 1993), which is carried out by means of the syntactic operation Copy.

(33) a. Peter [[$_V$ √TALK-CAUSE] [$_{SC}$ himself hoarse]]
      b. They [[$_V$ paint$_i$ [$_{SC}$ the wall √PAINT$_i$/red$_i$]]]

The strong resultative in (33a) involves conflation/compounding of a root and *v* (Merge), the compound being √TALK-CAUSE (much like √DANCE-V in (31)). The weak resultative in (33b) involves incorporation: the root (√PAINT) has moved to and incorporated into *v* (Copy). After this movement, a non-cognate object (*red*) is inserted. The non-cognate object is understood as being in a hyponymic relation to *paint* (similar to the way the PP *a l'habitació* (26/32) is treated as a cognate object of the abstract P conflated into the verb *entrar* 'to enter'; see footnote 226). In Mateu's terms:

---

[226] Following H&K's (2000) analysis of English particles like *up* in *We heated the soup up* as hyponymous and cognate complements of the verb (similar to *John danced a polka*), Mateu (2008) treats the PP *a l'habitació* as a case of P-cognation, i.e. it is also treated as a cognate complement of the verb.

(34) "(…) *weak* resultatives are possible when {Path/Result} is *incorporated* into *v* and the relevant tail of the movement is occupied by a coindexed non-cognate root, which is then semantically interpreted as hyponymous. That is, the Result root incorporated into v can be further specified by the resultative Adjective (…)."

In other words, in verb-framed languages like Romance the verb always involves directionality/resultativity, which can be further specified by means of an external phrase. Resultative constructions with non-directional verbs (i.e. strong resultatives) are not possible, which is the type of resultative found in English, for example.

Now comes the crucial question: can the analyses just described be applied to primary compounds? Mateu does indeed talk about compounding when deriving strong resultatives, but his understanding of compounding seems to be rather different from ours and no connection seems to be feasible. Mateu's use of the term 'compounding' when deriving (33a) refers to the operation of conflation, an operation that does not necessarily result in compounds as we defined them in chapter 2. To illustrate the point, on Mateu & Rigau's (2010) view, the denominal verb *siren,* as in *The horns sirened midday*, is derived by means of the compounding/conflation process (described above), the relevant compound being made up of a root and a null light verb, [v √*SIREN* V]. From this we conclude that their use of the term 'compound' and ours are incompatible with each other. The fact that in Mateu's analysis, strong resultatives (33a) (and complex path of motion constructions like *John danced into the room* (31)) are treated as verbal compounds which consist of a root and abstract verbs like 'cause' and 'go' is inconsequential for the types of compounds studied in chapter 2, since Mateu's analysis cannot be extended to them. That this conclusion seems to be correct is confirmed by Mateu's (2010) statement: "We show that Snyder's correlation between N-N compounds and complex resultatives cannot be maintained in his terms." In addition, in Mateu's analysis of strong resultatives, as we have just seen, the relevant compound is not formed by *talk* and *hoarse* (33a), as Snyder would claim, but by a root and a light verb. In short, Mateu's use of the term 'compounding' is different from ours, from which we conclude that his analysis cannot be applied to primary compounds. Having

said that, it is nevertheless not clear to us what evidence there is for postulating phonologically empty predicates like 'cause', 'go', 'become' and 'do' in syntax.[227]

This subsection concludes that the two analyses of resultatives briefly reviewed here (Kratzer's 2004 and Mateu's 2000, 2010) cannot be extended to cover primary compounds. There are empirical arguments against Kratzer's analysis and Mateu's conception of compounding is different from ours, which makes it impossible to apply his analysis to primary compounds. In short, it seems that resultative constructions and primary compounds cannot be treated as two instances of the same phenomenon.

3.2.2 On NN compounds: English vs. Catalan

From the previous subsections we have gathered that the availability of primary compounds, as exemplified by NN compounds, is not tied to the resultative construction and that their (non)existence cannot be treated in categorical terms (see Guevara & Scalise 2009 for related discussion). It is not the case that NN compounds are either present or absent in a language. Rather, there seems to be a continuum along which languages can be placed depending on the ease with which they can produce compounds. In this subsection it will be shown that English is placed at the pole of the scale where NN compounds can be freely formed, whereas Catalan is located at an intermediate point where NN compounds can be formed but within certain limits. Our proposal will be that mapping principles applying at the word level together with the morphological complexity of the compounding elements determine where in the continuum a language is to be located. Both factors are based on our belief that morphological features are the driving force behind phenomena like compounding.

Recall from chapter 1 (section 1.1) that we adopt a model of grammar (Jackendoff 1990, 1997, 2002, A&N 2004) in which a sentence has a semantic, syntactic and phonological representation. Each representation has its own primitives and may be non-isomorphic to the other representations. The three structures are connected by mapping principles. The same tripartite division is found at the word level: a word has a morphosyntactic, morphophonological and lexical-semantic representation. Each of them also has its own primitives and may be non-isomorphic to

---

[227] See also chapter 1 (section 1.1) for some criticisms of H&K's (1993, 1998, 2002) theory of argument structure, which Mateu assumes.

the other structures. The three representations of a word are also associated by mapping principles (see the model of grammar depicted in (11) in chapter 1). Only the morphosyntactic and morphophonological representations, along with the mapping between them, will be relevant in what follows. Following A&N (2004), a morpheme will be represented as MORPHEME in the morphosyntactic representation and as /morpheme/ in the morphophonological representation.

A&N (2004: 139-144) present three general mapping principles which are of interest here. They can be violated in subdomains of word formation in a language only if there is a trigger (e.g. conflicting demands between mapping principles). One mapping principle is linear correspondence: "crossing correspondences" between morphosyntactic and morphophonological representations are disfavoured (compare Sproat 1985: 82).

(35)   Linear Correspondence

      If   X is structurally external to Y,

              X is phonologically realized as /x/, and

              Y is phonologically realized as /y/

      then   /x/ is linearly external to /y/.

A second constraint is input correspondence, according to which an /affix/ takes as its host the phonological correspondent of the category that the AFFIX selects.

(36)   Input Correspondence

      If   an AFFIX selects (a category headed by) X,

              the AFFIX is phonologically realized as /affix/, and

              X is phonologically realized as /x/,

      Then   /affix/ takes /x/ as its host.

Let us see these two mapping principles at work. Consider the morphosyntactic structure in (37a), where the AFFIX is a /suffix/. Linear correspondence will favour the mapping from (a) to (b), while the input correspondence will favour the mapping to (c).[228]

---

[228] See Lardiere (1998: 289-290) for related discussion.

(37)   a. [[ₓ XY] AFFIX] ↔

b. [/x/ [/y/ /affix/]]

c. [[/x/ /affix/] /y/]

A&N (2004: 141, ex. 14)

The third mapping rule is quantitative correspondence (cf. Noyer 1993), according to which a single AFFIX cannot have more than one phonological representation.

(38)   Quantitative Correspondence

No element in the morphosyntax is spelled out more than once.

These three mapping principles will now be applied to compounds in Catalan and in English and we will consider how they can help explain why NN compounds are more productive in English than in Catalan.

Catalan NN compounds are left-headed and have two patterns of plural formation. Plural inflection can be placed on the head constituent only (39) or on both compounding elements (40).[229]

(39)   a. faldilla pantaló

skirt   trouser

'skort'

b. faldille+s pantaló

skirt-PL trouser

(40)   a. professor investigador

lecturer   researcher

'a lecturer who is also a researcher'

b. professor+s investigador+s

lecturer-PL researcher-PL

---

[229] A question that might arise regarding plural inflection being placed word-internally in Catalan compounds is whether the LIP is violated (Olsen 2001). It could be the case if plural inflection were syntactic but, in line with the framework adopted in this thesis, we treat it in the morphological component. Word structure can thus be uniformly generated in the same component.

In the case of (39), input correspondence and quantitative correspondence are satisfied because PLURAL is spelled out only once on the head. Precisely because /plural/ is on *faldilla*, it intervenes between the two elements, thus violating linear correspondence. Regarding (40), PLURAL is doubly realized: quantitative correspondence is clearly violated. Since the SUFFIX is realized on the second element, one could interpret that linear correspondence is satisfied because some spell-out of the SUFFIX is placed in the right position in the morphophonological representation: /plural/ is placed on *investigador*. If the same reasoning is extended to input correspondence, such a principle is not violated either: /plural/ is also placed on the head. In short, given the left-headedness of NN compounding in Catalan and the fact that PLURAL is spelled out by a /suffix/, there is no possible way in which all three principles can be satisfied. We believe that the unavoidable violation of a mapping principle restricts the productivity of compounding in Catalan, which in turn explains why the semantic patterns available in the language are more limited (see subsection 3.1.4 above), when compared to English.

By contrast, if NN compounds are right-headed and PLURAL is spelled out only once by a /suffix/, all three mapping principles can be satisfied. This is the case of English. Some examples are illustrative.

(41)   a. coffee table
        b. coffee tables

PLURAL is spelled out as a /suffix/ only once on the head, so input correspondence and quantitative correspondence are not violated. Linear correspondence is not violated either because nothing intervenes between the two elements. The fact that all mapping principles are satisfied makes the compounding process free of constraints and productive, which fits with the observation, repeatedly noted in the literature, that in English any two nouns can be combined and a plausible relation between them can be found.

In short, whether the mapping principles are satisfied or not seems to play a crucial role in determining the productivity of NN compounds. We will now turn to another factor which also seems to helps understand why NN compounding in Catalan is not as productive as in English.

We believe that the process of NN compounding is also sensitive to the morphological makeup of the elements to be compounded. Assume that there are two potential slots to be filled for each compounding element: the first is always filled while the second one is or is not filled depending on whether the element inserted in the first position is self-sufficient in the language. In a language like English, the second slot remains unfilled since the compounding element does not need the help of additional material (e.g. gender, thematic vowel) in order to be a well-formed word in the language. By contrast, in a language like Catalan, additional morphological material is necessary to conform to the morphological well-formedness conditions of the language. The head vs. non-head structure found in all cases of compounding, depicted in (86a) for English and (86b) for Catalan in chapter 2 and repeated below, will be given the structure in (42a) for English and (42b) for Catalan.

(86)   a.

non-head   head

b.

head   non-head
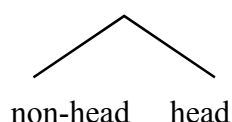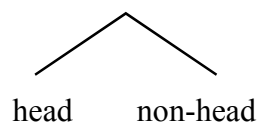
(42)   a.
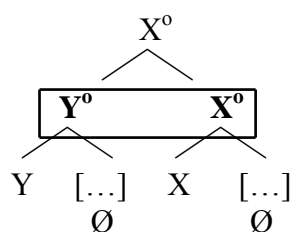
b.
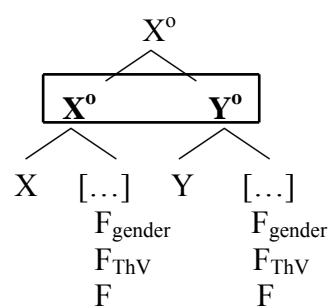
In English the compounding elements consist of bare lexemes (43), while in Catalan the compounding elements consist of lexemes which contain features encoding information about gender, the thematic vowel, among other features (44).[230] We assume that these features are inherently encoded in the noun and that they may be overtly or covertly

---

[230] The number and nature of the features are irrelevant to our concerns. What is crucial is that they are present in Catalan nouns while absent in English nouns.

expressed.[231,232] For example, we assume that there is a feature indicating gender in both (44a) and (44b) even though it is only visible in the latter case (by means of the vowel –*a*).

(43)   a. table leg

        b. clog dance

(44)   a. gos   llop

           dog-MASC wolf-MASC

           'wolf hound'

        b. gossa   lloba

           dog-FEM   wolf-FEM

We believe that the light vs. heavy nature of the compounding elements has repercussions on the productivity of the compounding process. Light compounding nodes facilitate while heavy compounding nodes hinder the ease with which compounds can be created. From this it follows that the compounding process can see whether the compounding nodes are simplex (as in English) or complex (as in Catalan) (see the boxed compound constituents in (42)).[233]

In the literature one can find other accounts to explain why compounding is a productive process in English while it is not so in Romance. Interestingly a number of them argue that the difference lies in the fact that the compound in Romance has a more complex structure than the compound in English. For example, Piera (1995: 306) uses

---

[231] We agree with authors like Varela (1989, 1990) when she says that gender in Spanish is an "inherent or substantial noun feature" (Varela 1989: 409). Mascaró (1986: 34-37) and Picallo (2008) defend the same view for Catalan. For example, consider what Picallo (2008: 49, fn. 6) says: "As is well known, all Catalan and Spanish common nouns must morphologically belong to one of two possible formal classes: the masculine or the feminine, henceforth [±fem]", "(…) we can very generally say that [+fem] prototypically surfaces as the suffix /a/ in both Catalan and Spanish. The value [-fem], the unmarked grammatical gender in both languages, mostly surfaces as the suffix /o/ in Spanish, and is phonologically null in Catalan. (…)".

[232] Interestingly, in the same way that inherently/lexically encoded features like gender have repercussions on the compounding process, Horrocks & Stavrou (2007) report that inherently/lexically encoded features in the verb system also have consequences for a syntactic process. More precisely, the presence of grammaticalized aspect in Greek (verbal morphology necessarily bears aspect marking) is tied to the unavailability of secondary resultative predication.

[233] As the conclusion to this subsection indicates, our proposal needs further study. As it stands now, it would not be able to explain why Dutch NN compounding is as productive as it is in English given that Dutch nouns have gender, like Catalan. What may be at stake here is the fact that gender in nouns is never formally expressed in Dutch, unlike Catalan (see footnote 231). I am grateful to Peter Ackema for this observation.

the constraint in (45) to explain the contrast between NN compounds in English and in Spanish.

(45)   "A double bracket at the edge of a word blocks adjunction of a word."

Details aside, the double bracket in the Spanish compound arises as a consequence of having a word marker (WM), absent in English. The WM is postulated for Spanish nouns, adjectives and adverbs. Compare the structure of *dog* vs. *perro* 'dog' in (46) and the minimal shape of English vs. Spanish compounds in (47).[234]

(46)   a. [$_{+N...}$ dog]
       b. [$_{+N...}$ [perr(o)] WM]

(47)   a. [$_{+N...}$ [$_{+N...}$ police] dog]
       b. [$_{+N...}$ [[perr(o)] WM] [[policí(a)] WM]]

The double bracket arises because there is extra material, morphological in nature, to be accommodated within the Spanish compound. Similar remarks are found in relation to Italian compounds. Consider the following quote from Scalise & Bisetto (2009: 36):

(48)   "In an Italian compound such as *capostazione* 'station master', the words *capo* and *stazione* are not 'roots' even though each of them can be described as being formed by a root (*cap-* and *stazion-*) plus a morpheme (*-o* and *–e*, respectively) indicating the inflectional class to which the lexemes belong."

In short, our proposal bears some resemblance to other proposals available in the literature in that we all attribute the difference between English and Romance compounding to the morphological complexity of the compounding elements.

---

[234] On Piera's (1995) account the presence of the double bracket in Spanish compounds explains why they cannot be right-headed (i) and why they cannot be recursive (ii)

       (i)   police dog vs. *policía perro (lit. police dog)
       (ii)   pet police dog vs. *perro policía mascota (lit dog police pet)

See the original work for details.

We are aware that the proposal made in this subsection needs further elaboration. At this stage our account does not explain why some compounds and some readings are not possible in Catalan. For example, why should *pot cuc* (can worm) 'a can which has the shape of a worm, a can which can be changed into a worm, etc.' not allow the goal/locative reading which its English counterpart has? Consider the definition of *worm can* given by Snyder (2001: 338): "a can in which fishing bait is stored". In addition, it is not clear in which way 'heavy' compounding nodes (as opposed to 'light' compounding nodes) impede the productivity of NN compounding. We believe that a cross-linguistic comparison would help elucidate these unclear questions, but we leave this for future research. For now we conclude that the morphological pattern of the compound seems to determine whether NN compounding will be productive in the language (NN compounds are freely formed in English while they are subject to more restrictions in Catalan) but does not say anything about the (un)availability of resultatives. We have seen that the morphological pattern must include information about plural formation and must also know if features like gender need to be present in the compounding constituents. From the discussion above, it follows that a language will be located in a higher or lower position in the continuum measuring the productivity of NN compounds depending on the type of morphological pattern allowed in the language.

3.3 Conclusions

This chapter started by presenting Snyder's (2001) Compounding Parameter with the subsequent changes it has undergone. The alleged dependence relation of resultatives on the availability of primary compounds (NN compounding) was examined in English (subsection 3.1.1), in Catalan (subsection 3.1.2) and in other language groups (subsections 3.1.3 and 3.1.4). After discussing some controversial points regarding the TCP (subsection 3.1.4) we concluded that the alleged correlation between NN compounding and strong resultatives is not as strong as is claimed by Snyder: the implicational relationship from complex predicates to the positive setting of the TCP as well as the interdependence among the complex predicates (which are argued to form a natural class by Snyder (1995f)) are threatened.

In the second part of the chapter two syntactic accounts of adjectival resultatives were presented: Kratzer's (2005) and Mateu's (2000, 2010). The possibility of

extending their analyses to primary compounds was considered but we concluded that such an extension is not possible. In addition, neither account is designed to explain why NN compounds are more easily formable in English than in Catalan if they exist in the two languages, an issue which was taken up in subsection 3.2.2 where it was concluded that the ease with which NN compounds can be created seems to depend on mapping principles applying at the word level and on the morphological complexity of the compounding elements.

4. Conclusions

This chapter presents the main conclusions of the present work in a compact summary form as well as some questions which need to be addressed in future research.

This thesis has considered a type of complex word-formation - compounding - and its relation to the morphology-syntax interface. In the first chapter, arguments were given for the plausibility of a model of grammar in which word syntax (referred to as morphology) and phrasal syntax (referred to as syntax) are two distinct modules within a bigger syntactic module (cf. Jackendoff 1990, 1997, 2002, A&N 2004) (section 1.1). In addition, some evidence was provided for the generation of complex words, compounds included, in the morphological subcomponent (section 1.2). Consider (1).

(1)   SYNTAX
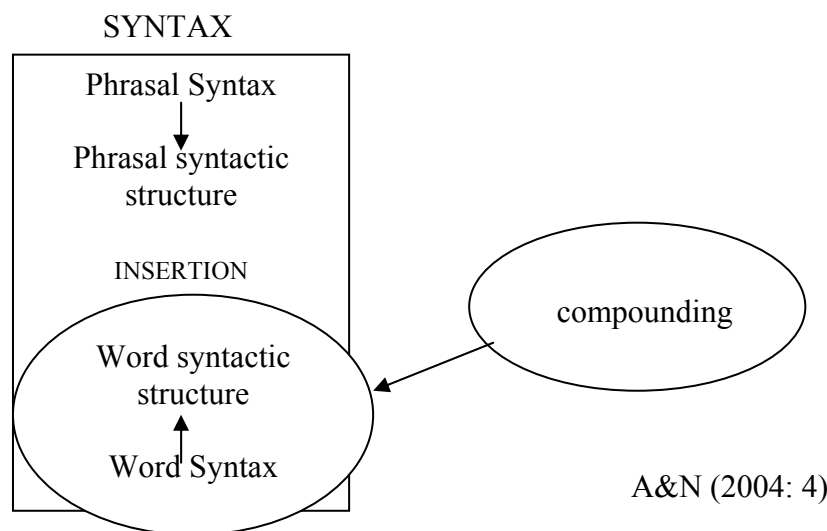
A&N (2004: 4)

A&N's (2004) morphologically-based account of compounding, based on their competition model between syntax and morphology, was shown to be able to satisfactorily account for compounding in English and Romance (Catalan and Spanish), provided a semantic constraint assumed in their theory was refined (section 1.3). By contrast, Harley's (2004, 2008b) syntactically-based account of compounding, based on DM (Halle & Marantz 1993, Marantz 1997a, b, 2001, 2007, Harley & Noyer 1999, Embick & Noyer 2007, a.o.), a model in which syntax (the only generative component) is responsible for both word and phrase structure, was shown to have some problems for the correct generation of compounds in English (section 1.4).

In the second chapter the existence of heads in morphology was discussed, confirmed (contra Zwicky 1985, Bauer 1990, and Anderson 1992) and shown to play a fundamental role in the classification of compounds in English and Catalan (section 2.1). Afterwards (subsection 2.2.1) the nature of the compounding elements in English and Catalan was identified. We concluded that English compounds can consist of (i) a root, a lexeme or a phrase in the non-head position, and (ii) a root or a lexeme in the head position. Regarding Catalan compounds, they can consist of (i) a root, a stem, a lexeme or a phrase in the non-head position, and (ii) a root, a stem or a lexeme in the head position. In both languages, roots are only present in the case of neo-classical compounds. Then (subsection 2.2.2), a brief overview of several types of compound classifications (Lees 1960, Hatcher 1960, Levi 1978, Downing 1977, Bauer 1983, 2003, Booij 2005, Plag 2003) was provided. After concluding that none of them was satisfactory enough, what looked like the most promising classification of compounds available in the literature was explored: B&S's (2005) classification (see below). It provides three macro-types of compounds, which are based on the grammatical relation between the two constituents: SUB, ATR and CRD. Each macro-type is in turn divided into two subtypes: endocentric vs. exocentric.

(2)

Compounds

Subordinate (SUB)   Attributive (ATR)   Coordinate (CRD)

endo.	exo.	endo.	exo.	endo.	exo.
apple cake	kill joy	ape man	white collar	actor author	mind brain
taxi driver	cut throat	key word	pale face	dancer singer	north east
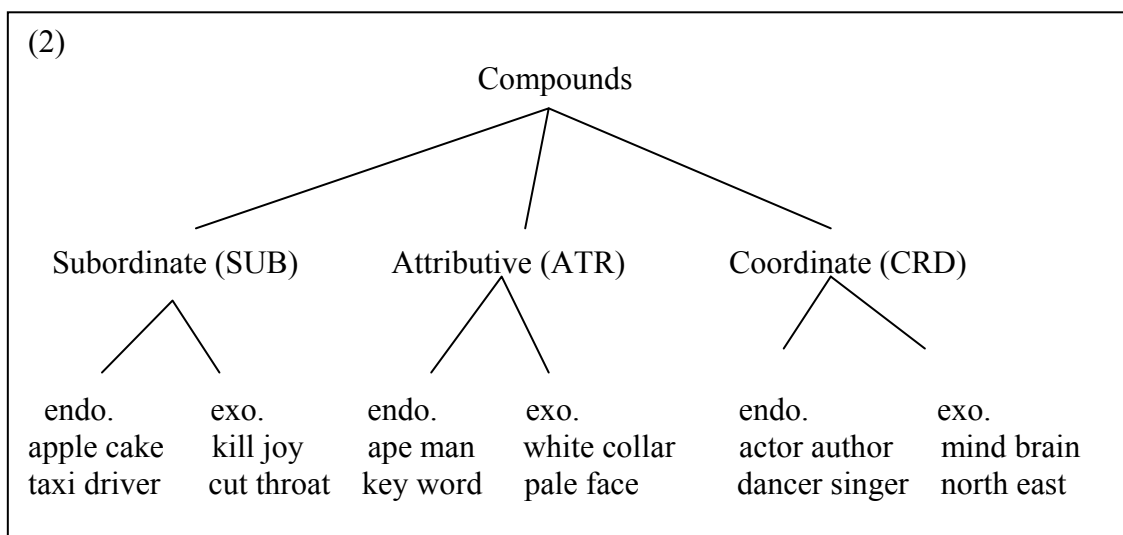
B&S (2005: 326)

B&S acknowledge that their classification contains rough subdivisions and that adding another layer of analysis to their two levels would refine it, a task which was partially carried out in the following section (section 2.3). That is, we incorporated the category of the input compounding elements and the category of the resulting compound into the

original scheme of B&S, thus creating further subdivisions in their classification. The resulting scheme was used when carrying out an exhaustive study of the compounds available in English and Catalan (subsections 2.3.1 and 2.3.2 respectively). Although B&S's classification was our starting point, we departed from it substantially. We reduced B&S's three macro-types of compounds to one type. Our proposal was that all compounds are based on a head vs. non-head relation, from which the different interpretations arise (SUB, ATR). We suggested that the (context and) semantics of the compounding elements determines whether the non-head has to be interpreted as a kind of attribute or complement to the head. The existence of CRD compounds and exocentric compounds was denied: we suggested that what are generally called CRD compounds are cases of asyndetic syntactic coordination (not compounds) and that all compounds are endocentric. We concluded that productive English compounds have the structure given in (3a) while Catalan compounding is mainly characterized by the structure given in (3b).

(3)　　a.　　　　　　　　　　　　　　　　　b.

　　　　　non-head　head　　　　　　　　　head　non-head

In the third chapter, Snyder's Compounding Parameter was presented (Snyder 1995, 1996, 2001, 2002), along with some revisions it has undergone (Beck & Snyder 2001a, Snyder et al. 2001, Roeper et al. 2002, Roeper & Snyder 2005, Snyder 2005). After identifying which complex predicates were relevant to the [TCP] (e.g. strong resultatives, verb-particle constructions including non-directional verbs), the workings of the [TCP] were considered in a few languages. We concluded that a strict application of the compounding/complex-predicate parameter cannot be maintained. In addition to the apparently exceptional behaviour of Basque in being a [+TCP] language and yet not having strong adjectival resultatives, there are a few other languages (e.g. Catalan, Greek, Japanese) which behave in the same way. Snyder's alleged dependence of complex predicates on NN compounding was also questioned: Latin and Slavic languages like Russian have resultatives but no NN compounding. Additionally, Snyder's claim that complex predicate constructions are learned as a block had to be discarded in view of languages like Catalan, Greek, Russian and Latin that allow only some of them. In short, the validity of the [TCP] was denied (section 3.1).

The second part of the chapter (section 3.2) considered whether there is a real connection between resultatives and compounding. To this end, two syntactic analyses of resultatives (Kratzer's 2005 and Mateu's 2000, 2010) were briefly reviewed. The possibility of extending their analyses to primary compounds was considered but we concluded that such an extension is not possible. Compounding and resultative constructions seem to be two different phenomena. Finally, we addressed the question of why in some languages - like Catalan - NN compounds are productive, albeit to a lesser degree than NN compounds in a language like English. We suggested that the answer could be found in some mapping principles applying at the word level and in the morphological complexity of the compounding elements (subsection 3.2.2).

In the course of answering the central questions of this thesis, other questions have been mentioned in passing while others have only been briefly tackled: e.g. DM's claim that roots are acategorial (subsection 1.4.3), criteria for distinguishing compounds from non-compounds like appositions (section 2.3), and the role of productivity in compounding (sections 2.3 and 3.2.2) are just a few of them. While these and other questions certainly merit further study, they fall outside the scope of the present study.

At this point we feel it seems appropriate to stop, even if things are just beginning to look exciting.

References

Acedo-Matellán, V. 2009. "Adjectival resultatives cross-linguistically: a morphoponological account". Talk given at ConSOLE XVII, University of Nova Gorica, Nova Gorica.

Ackema, P. 1999a. *Issues in Morphosyntax*. Amsterdam: John Benjamins.

Ackema, P. 1999b. "The nonuniform structure of Dutch N-V compounds". In Booij, G. and van Marle, J. (eds.). *Yearbook of Morphology* 1998. Dordrecht: Kluwer. pp. 127-158.

Ackema, P. and Neeleman, A. 2002. "Morphological selection and representational modularity". In Booij, G. and van Marle, J. (eds.). *Yearbook of Morphology 2001*. Dordrecht: Kluwer. pp: 1-51.

Ackema, P. and Neeleman, A. 2003. "Syntactic Atomicity". *Journal of Comparative Germanic Linguistics* 6: 93-128.

Ackema, P. and Neeleman, A. 2004. *Beyond Morphology*: *Interface Conditions on Word Formation*. Oxford: Oxford University Press.

Ackema, P. and Neeleman, A. 2007. "Morphology does not equal Syntax" In Ramchand, G. and Reiss, C. (eds). *The Handbook of Linguistic Interfaces*. Oxford: Oxford University Press. pp. 325-352.

Ackema, P. and Neeleman, A. 2008. "Synthetic compounds are derived root compounds: Evidence from creolization". Talk given at the CompoNet Congress on Compounding, University of Bologna, Bologna.

Ackerman, F. and LeSourd, P. 1997. "Toward a Lexical Representation of Phrasal Predicates. In Alsina, A., Bresnan, J. and Sells, P. (eds.). *Complex Predicates*. Stanford: CSLI Publications. pp. 67-106.

Adams, V. 1973. *An introduction to modern English word formation*. London: Longman.

Adams, V. 2001. *Complex Words in English*. Harlow: Longman.

Adelman, A. 2002. *On the status of N-V verbal compounds in Catalan*. Senior Honor Thesis, Swarthmore College.

Alexiadou, A. 2003. "Some notes on the structure of alienable and inalienable possessors". In Coene, M. and d'Hulst, Y. (eds.). *From NP to DP (vol.2): The expression of possession in noun phrases*. Amsterdam: John Benjamins. pp. 167-188.

Allen, M. R. 1978. *Morphological Investigations*. PhD dissertation, University of Connecticut.

Amritavalli, R. and Jayaseelan, K. A. 2003. "The genesis of syntactic categories and parametric variation". Talk given at the 4th Asian GLOW Colloquium, Seoul.

Anagnostopoulou, E. 1994. *Clitic dependencies in Modern Greek*. PhD dissertation, University of Salzburg.

Anagnostopoulou, E. 2003. *The Syntax of Ditransitives: Evidence from Clitics*. Berlin: Mouton de Gruyter.

Anderson, S. 1982. "Where's Morphology?" *Linguistic Inquiry* 13: 571-612.

Anderson, S. 1992. *A-Morphous Morphology*. Cambridge: Cambridge University Press.

Anderson, A., Brown, L., Gaby, A. and Lecarme, J. 2006. "Life on the Edge: there's Morphology there after all". *Lingue e Linguaggio* 1 [Special Issue on Lexical Integrity]: 33-47.

Archangeli, D. and Pulleyblank, D. 1994. *Grounded phonology*. Cambridge, Mass.: MIT Press.

Aronoff, M. 1976. *Word Formation in Generative Grammar*. Cambridge, Mass.: MIT Press.

Aronoff, M. 1994a. *Morphology by itself: stems and inflectional classes*. Cambridge, Mass.: MIT Press.

Aronoff, M. 1994b. "Blocking". In Asher, R. E. and Simpson, J. M. Y. (eds.). *The Encyclopedia of Language and Linguistics*, vol.1. Oxford: Pergamon Press.

Baayen, H. 1992. "Quantitative aspects of morphological productivity". In Booij, G. and van Marle, J. (eds.). *Yearbook of Morphology* 1991. Dordrecht: Kluwer. pp. 109-149.

Badia i Margarit, A. M. 1962. *Gramática catalana*. Madrid: Gredos.

Baker, M. 1985. "The Mirror Principle and Morphosyntactic Explanation". *Linguistic Inquiry* 16: 373-415.

Baker, M. 1988. *Incorporation: A Theory of Grammatical Function Change*. Chicago & London: University of Chicago Press.

Baker, M. 1993. "A review of *A-Morphous Morphology* by Stephen Anderson". *Language* 69: 587-590.

Baker, M. 1996. *The Polysynthesis Parameter*. New York: Oxford University Press.

Baker, M. 1997. "Thematic Roles and Syntactic Structure". In Haegeman, L. (ed.). *Elements of Grammar*. Dordrecht: Kluwer. pp. 73-137.

Baker, M. 2003. *Lexical categories: verbs, nouns, and adjectives*. Cambridge: Cambridge University Press.

Baker, M. 2009. "Is head movement still needed for noun incorporation?" *Lingua* 119: 148-165.

Bally, C. 1950. *Linguistique générale et linguistique francaise*. Berne : A. Francke. (3rd edition)

Bartra, A. and Suñer, A. 1992. "Functional Projections Meet Adverbs". *Catalan Working Papers in Linguistics* 1992: 45-85.

Battus. 1985. *Opperlandse taal- en letterkunde*. Amsterdam: Querido.

Bauer, L. 1983. *English Word-Formation*. Cambridge: Cambridge University Press.

Bauer, L. 1990. "Be-heading the word". *Journal of Linguistics* 26: 1-31.

Bauer, L. 1998. "When is a sequence of two nouns a compound in English?" *English Language and Linguistics* 2: 65-86.

Bauer, L. 2001. "Compounding". In Haspelmath, M., König, E., Oesterreicher, W. and Raible, W. (eds.). *Language Typology and Language Universals*, vol. 1. Berlin: Walter de Gruyter. pp. 695-707.

Bauer, L. 2003. *Introducing Linguistic Morphology*. Edinburgh: Edinburgh University Press.

Bauer, L. 2004. "Adjectives, Compounds and Words". *Nordic Journal of English Studies* 3: 7-22.

Bauer, L. 2008. "Dvandva". *Word Structure* 1: 1-20.

Bauer, L. and Renouf, A. 2001. "A Corpus-Based Study of Compounding in English". *Journal of English Linguistics* 29: 101-123.

Bauer, L. and Valera, S. 2005. *Approaches to conversion/zero-derivation*. Münster: Waxmann Verlag.

Beard, R. 1995. *Lexeme-Morpheme Base Morphology*. Albany: SUNY Albany Press.

Beck, S. and Snyder, W. 2001a. "Complex Predicates and Goal PP's: Evidence for a Semantic Parameter". In Do, A. H.-J., Domínguez, L. and Johansen, A. (eds.). *BUCLD 25: Proceedings of the 25th Annual Boston University Conference on Language Development* (vol. 1). Somerville, Mass.: Cascadilla Press. pp. 114-122.

Beck, S. and Snyder, W. 2001b. "The resultative parameter and restitutive again". In Frey, C. and Sternefeld, W. (eds.). Auditur vox sapientiae: a Festschrift for Arnim von Stechow. Berlin: Akademie Verlag. pp. 81-95.

Benczes, R. 2006. *Creative compounding in English: the semantics of metaphorical and metonymical noun-noun combinations*. Amsterdam: Benjamins.

Benveniste, É. 1974. *Problèmes de linguistique générale* II. Paris: Gallimard.

Berg, T. 1998. "The (in)compatibility of morpheme orders and lexical categories and its historical implications". *English Language and Linguistics* 2: 245-262.

Bisetto, A. 2004. "Composizione con elementi italiani". In Grossmann, M. and Rainer, F. (eds.). *La formazione delle parole in italiano*. Tübingen: Niemeyer. pp. 33-55.

Bisetto, A. and Melloni, C. 2008. "Parasynthetic Compounds". Talk given at the CompoNet Congress on Compounding, University of Bologna, Bologna.

Bisetto, A. and Scalise. S. 1999. "Compounding: morphology and/or syntax". In Mereu, L. (ed.). *Boundaries of Morpholgy and Syntax*. Amsterdam/Philadelphia: John Benjamins. pp. 31-48.

Bisetto, A. and Scalise, S. 2005. "Classification of compounds". *Lingue e Linguaggio* 2: 319-332.

Bloomfield, L. 1933. *Language*. London: George Allen & Unwin.

Boertien, H. 1997. "Left-headed compound prepositions". *Linguistic Inquiry* 28: 689-697.

Bok-Bennema, R. and Kampers-Manhe, B. 2006. "Taking a closer look at Romance VN compounds". In Nishida, C. and Montreuil, J.-P. (eds.). *New Perspectives on Romance Linguistics*, vol. 1. Amsterdam Philadelphia: John Benjamins. pp. 13-26.

Bonet, E. 1991. *Morphology after syntax. Pronominal clitics in Romance*. PhD dissertation, MIT.

Bonet, E. 2007. "Realització de trets morfosintàctics i Morfologia Distribuïda: una introducció al model". Talk given at CLT of Universitat Autònoma de Barcelona, Barcelona.

Booij, G. 1985. "Coordination reduction in complex words: a case for prosodic phonology". In van der Hulst, H. and Smith, N. (eds.). *Advances in nonlinear phonology*. Dordrecht: Foris. pp. 143-160.

Booij, G. 1988. "The Relation between Inheritance and Argument Linking: Deverbal Nouns in Dutch". In Everaert, M., Evers, A., Huybregts, R. and Trommelen, M. (eds.). *Morphology and Modularity*. Dordrecht: Foris. pp. 57-73.

Booij, G. 2005. *The Grammar of Words*. Oxford: Oxford University Press.

Booij, G. 2007. "Construction Morphology and the Lexicon". In Montermini, F., Boyé, G. and Hathout, N. (eds). *Selected Proceedings of the 5th Décembrettes: Morphology in Toulouse*. Somerville, Mass.: Cascadilla Proceedings Project. pp. 34-44.
http://www.lingref.com/cpp/decemb/5/index.html

Bordelois, I. 1993. "Afijación y estructura temática: *-Da* en español". In Varela, S. (ed.). *La formación de palabras*. Madrid: Taurus. pp. 162-179.

Borer, H. 1984. *Parametric Syntax*. Dordrecht: Foris.

Borer, H. 1989. "On the Morphological Parallelism between Compounds and Constructs". In Booij, G. and van Marle, J. (eds.). *Yearbook of Morphology* 1. Dordrecht: Foris. pp. 45-65.

Borer, H. 1998. "Morphology and Syntax". In Spencer, A. and Zwicky, A. M. (eds.). *The Handbook of Morphology*. Oxford: Blackwell. pp. 151-190.

Borer, H. 2003. "Exo-Skeletal vs. Endo-Skeletal Explanations: Syntactic Projections and the Lexicon". In Moore, J. and Polinsky, M. (eds) *The Nature of Explanation in Linguistic Theory*. Stanford: CSLI Publications. pp. 31-67.

Borer, H. 2008. "Compounds: the view from Hebrew". Ms. [The final version can be found in: Lieber, R. and Štekauer, P. (eds.). 2009. *The Oxford Handbook of Compounding*. Oxford: OUP. pp. 491-511.]

Borer, H. 2009. "Roots and categories". Talk given at the 19th Colloquium on Generative Grammar, University of the Basque Country, Vitoria-Gasteiz.

Bosque, I. 1989. *Las categorías gramaticales. Relaciones y diferencias*. Madrid: Síntesis.

Botha, R. 1981. "A Base Rule Theory of Afrikaans Synthetic Compounding". In Moortgat, M., van der Hulst, H. and Hoekstra, T. (eds.). *The Scope of Lexical Rules*. Dordrecht: Foris. pp. 1-77.

Botha, R. 1984. *Morphological Mechanisms*. Oxford, UK: Pergamon.

Bouchard, D. 1995. *The Semantics of Syntax: A Minimalist Approach to Grammar*. Chicago & London: University of Chicago Press.

Branchadell, A. 1992. *A Study of Lexical and Non-Lexical Datives*. PhD dissertation, Universitat Autònoma de Barcelona.

Brattico, P. 2005. "A Category-free Model of Finnish Derivational Morphology". *SKY Journal of Linguistics* 18: 7-45.

Brekle, H. E. 1970. *Generative Satzsemantik und transformationelle Syntax im System der Englischen Nominalkomposition*. Munich: Fink.

Bresnan, J. and Mchombo, S. 1995. "The Lexical Integrity Principle: Evidence from Bantu". *Natural Language and Linguistic Theory* 13: 181-254.

Brunelli, M. 2003. *Un grup especial de verbs compostos en català*. Tesi de Laurea, Università Ca' Foscari.

Buenafuentes, C. 2001-2002. "Procesos de gramaticalización en el estudio de los compuestos del español: el caso del elemento mal/a(-)". *Anuari de Filologia*, volum XXIII-XXIV, secció F, 11-12: 17-29.

Buenafuentes, C. 2007. *Procesos de gramaticalización y lexicalización en la formación de compuestos en español*. PhD dissertation, Universitat Autònoma de Barcelona.

Cabré, M. T. 1988. "La prefixació en català". In Staczeck, J. J. (ed.). *On Spanish, Portuguese and Catalan Linguistics*. Washington: Georgetown Univeristy Press. pp. 47-63.

Cabré, M. T. 1994. *A l'entorn de la paraula*, vol. 2. València: Servei de Publicacions de la Universitat de València.

Cabré, M. T. 2002. "La derivació". In Solà, J., Lloret, R.-M., Mascaró, J. and Pérez Saldanya, M. (eds.). *Gramàtica del català contemporani*. Barcelona: Empúries. pp. 731-775

Cabré, M. T. and Rigau, G. 1986. *Lexicologia i Semàntica.* Barcelona: Enciclopèdia Catalana.

CALD: Cambridge Advanced Learner's Dictionary http://dictionary.cambridge.org/

Carnie, A. 2000. "On the notions XP and X^0". *Syntax* 3: 59-106.

Carstairs-McCarthy, A. 1992. *Current Morphology*. London: Routledge.

Carstairs-McCarthy. A. 1993. "Morphology without word-internal constituents: a review of Stephen R. Anderson's *A-Morphous Morphology*". In Booij, G. and van Marle, J. (eds.). *Yearbook of Morphology* 1992. Dordrecht: Kluwer. pp. 209-233.

Carstairs-McCarthy. A. 2002. *An Introduction to English Morphology*. Edinburgh: Edinburgh University Press.

Ceccagno, A. and Basciano, B. 2007. "Compound headedness in Chinese: an analysis of neologisms". *Morphology* 17: 207-231.

Chomsky, N. 1957. *Syntactic Structures*. The Hague: Mouton.

Chomsky, N. 1965. *Aspects of the Theory of Syntax*. Cambridge, Mass.: MIT.

Chomsky, N. 1970. "Remarks on Nominalization." In Jacobs R. and Rosenbaum, P. (eds.). *Readings in English Transformational Grammar*. Waltham, Mass.: Ginn and co. pp. 184-221.

Chomsky, N. 1981. *Lectures on government and binding*. Dordrecht: Foris.

Chomsky, N. 1993. "A Minimalist Program for Linguistic Theory". In Hale, K. and Keyser, S. J. (eds.). *A view from Building 20*: *Essays in Linguistics in Honor of Sylvian Bromberger*. Cambridge, Mass.: MIT Press. pp. 1-52. [reprinted in Chomsky 1995a]

Chomsky, N. 1995a. *The Minimalist Program*. Cambridge, Mass.: MIT Press.

Chomsky, N. 1995b. "Bare Phrase Structure". In Webelhuth, G. (ed.). *Government and Binding theory and the Minimalist Program*. London: Blackwell. pp. 383-439.

Chomsky, N. 2000. "Minimalist Inquiries: The Framework". In Martin, R., Michaels, D. and Uriagereka, J. (eds.). *Step by Step. Essays on Minimalist Syntax in Honour of Howard Lasnik*. Cambridge, Mass.: MIT Press. pp. 89-155.

Chomsky, N. 2001. "Derivation by Phase". In Kenstowicz, M. (ed.). *Ken Hale: A Life in Language*. Cambridge, Mass.: MIT Press. pp. 1-52.

Chomsky, N. 2004. "Beyond Explanatory Adequacy". In Belletti, A. (ed.). *Structures and Beyond. The Cartography of Syntactic Structures*, vol.3. Oxford, NY: Oxford University Press. pp. 104-131.

Chomsky, N. 2005. "On Phases". Ms. MIT. [A published version can be found in: Freidin, R., Otero, C. P. and Zubizarreta, M. L. (eds.). 2008. *Foundational Issues in Linguistic Theory*. Cambridge, Mass.: MIT Press.]

Chomsky, N. 2007. "Approaching UG from below". In Sauerland, U. and Gärtner, H-M. (eds.). *Interfaces + Recursion = Language? Chomsky's minimalism and the view from syntax-semantics*. Berlin: Mouton de Gruyter. pp. 1-30. [cited by manuscript version]

Chomsky, N. and Halle, M. 1968. *The sound pattern of English*. New York: Harper & Row.

Cinque, G. 1993. "A Null Theory of Phrase and Compound Stress". *Linguistic Inquiry* 24: 239-297.

Clark, E. V. 1993. *The Lexicon in acquisition*. Cambridge/New York: Cambridge University Press.

Clark, E. V. 2003. *First language acquisition*. Cambridge: Cambridge University Press.

Clark, E. V. and Hecht, B. 1982. "Learning to Coin Agent and Instrument Nouns". *Cognition* 12: 1-24.

Clark, E. V., Hecht, B. and Mulford, R. 1986. "Acquiring Complex Compounds: Affixes and Word Order in English". *Linguistics* 24: 7-29.

Clements, J. C. 1992. "Lexical category hierarchy and 'head of compound' in Spanish". In Laeufer, C. and Morgan, T. A. (eds.). Theoretical *Analyses in Romance Linguistics*. Amsterdam/Philadephia: John Benjamins. pp. 151-166.

Contreras, H. 1985. "Spanish Exocentric Compounds". In Nuessel, F. H. (ed.). *Current Issues in Hispanic Phonology and Morphology*. Bloomington: IUCL. pp. 14-27.

Davidson, D. 1967. "The logical form of action sentences". In Roscher, N. (ed.). *The Logic of Decision and Action*. Pittsburg: University of Pittsburg. pp. 81-95.

Delfitto, D. and Melloni, C. 2008. "Compounds don't come easy". Talk given at the Workshop Words don't come easy, University of Verona, Verona.

Demonte, V. 1992. "Temporal and aspectual constraints on predicative adjective phrases". In Campos, H. and Martínez-Gil, F. (eds.). *Current Studies in Spanish Linguistics*. Washington, D.C.: Georgetown University Press. pp. 165-200.

Demonte, V. 1994a. "Datives in Spanish". *University of Venice Working Papers in Linguistics* 4: 71-96.

Demonte, V. 1994b. "On Certain Asymmetries between DOs and IOs". In Cinque, G., Koster, J., Pollock, J.-Y., Rizzi, L. and Zanuttini, R. (eds.) *Paths towards universal grammar: studies in honor of Richard S. Kayne*. Washington, D.C.: Georgetown University Press. pp. 111-120.

Demonte, V. 1995. "Dative alternation in Spanish". *Probus* 7: 5-30.

Demonte, V. 1999. "Semántica composicional y gramática: los adjetivos en la interficie léxico-sintaxis". *Revista Española de Lingüística* 29: 283-316.

Demonte, V. and Masullo, P. J. 1999. "La predicación: Los complementos predicativos". In Bosque, I. and Demonte, V. (eds.). *Gramática Descriptiva de la Lengua Española*. Madrid: Espasa Calpe. pp. 2461-2523.

Devereux, B. and Costello, F. 2007. "Learning to interpret novel noun-noun compounds: evidence from a category learning experiment". *Proceedings of the Workshop on Cognitive Aspects of Computational Language Acquisition*: 89-96.

DIEC: *Diccionari Institut d'Estudis Catalans* http://dlc.iec.cat/

Dikken, M. den. 1992. *Particles*. Dordrecht: Holland Institute of Generative Linguistics.

Di Sciullo, A. M. 1997. "Prefixed-verbs and Adjunct Identification". In Di Sciullo, A. M. (ed.). *Projections and Interface Conditions*. New York / Oxford: Oxford University Press. pp. 52-73.

Di Sciullo, A. M. 2005. *Asymmetry in Morphology*. Cambridge, Mass.: MIT Press.

Di Sciullo, A. M. 2007. "Words, Phases and Compositionality". Talk given at XXXIII Incontro di Grammatica Generativa, University of Bologna, Bologna.

Di Sciullo, A. M. and Williams, E. 1987. *On the Definition of Word*. Cambridge, Mass.: MIT Press.

Dijk, S. 1997. *Noun Incorporation in Frisian*. PhD dissertation, University of Groningen. Ljouwert: Fryske Akademy.

DOD: *Datasegment Online Dictionary* http://onlinedictionary.datasegment.com/

Don, J. 1993. *Morphological Conversion*. PhD dissertation, Utrecht University.

Don, J. 2003. "Roots, Deverbal Nouns and Denominal Verbs". In Booij, G., Guevara, E., Ralli, A., Sgroi, S. and Scalise, S. (eds.). *Morphology and Linguistic Typology. On-line proceedings of the Fourth Mediterranean Morphology Meeting (MMM4)*. University of Bologna, 2005. pp. 91-103. http://morbo.lingue.unibo.it/mmm/

Don, J. 2004. "Categories in the lexicon". *Linguistics* 42: 931-956.

Don, J., Trommelen, M. and Zonneveld, W. 2000. "Conversion and category indeterminacy". In Booij, G., Lehmann, C. and Mugdan, J. (eds.). *Morphology: An International Handbook on Inflection and Word-Formation*. Berlin: Walter de Gruyter. pp. 943-952.

Downing, P. 1977. "On the Creation and Use of English Compound Nouns". *Language* 53: 810-842.

Dowty, D. R., Wall, R. E. and Peters, S. 1981. *Introduction to Montague Semantics*. Dordrecht: Reidel.

DRAE: *Diccionario Real Academia Española* http://www.rae.es (22nd edition)

Duarte, C. and Alsina, A. 1986. *Gramàtica històrica del català*, vol. 2. Barcelona: Curial.

Dubinsky, S. and Simango, S. R. 1996. "Passive and stative in Chichewa: Evidence from modular distinctions in grammar". *Language* 72: 749-781.

Embick, D. and Marantz, A. 2008. "Architecture and blocking". *Linguistic Inquiry* 39: 1-53.

Embick, D. and Noyer, R. 2007. "Distributed Morphology and the Syntax-Morphology Interface". In Ramchand, G. and Reiss, C. (eds.). *The Oxford Handbook of Linguistic Interfaces*. Oxford: Oxford University Press. pp. 289-324.

Emonds, J. E. 1976. *A transformational approach to English syntax*. London: Academic Press.

Emonds, J. 2006. "How much should we Distribute Morphology". *Theoretical and Applied Linguistics at Kobe Shoin* 9: 1-16.

Fabb, N. 1984. *Syntactic Affixation*. PhD dissertation, MIT.

Fabb, N. 1998. "Compounding". In Spencer, A. and Zwicky, A. M. (eds.). *The Handbook of Morphology*. Oxford: Blackwell. pp. 66-83.

Fabra, P. 1956. *Gramàtica catalana*. Barcelona: Teide. (9th edition)

Fábregas, A. 2006. "Historia de un problema: la morfología desde 1985 hasta hoy". In Felíu, E., Gil Laforga, I. and Torres, M. (eds.). *La Morfología a debate*. Jaén: Universidad de Jaén. pp. 63-92.

Fábregas, A., Felíu, E. and Varela, S. 2006. "The Lexical Integrity Hypothesis and Morphological Local Domains". *Lingue e Linguaggio* 1 [Special Issue on Lexical Integrity]: 83-104.

Fábregas, A. and Scalise, S. 2008. "Exocentricity in Compounding: How extended and how problematic". Talk given at the Workshop Words don't come easy, University of Verona, Verona.

Farrell, P. 2005. "English verb-preposition constructions: Constituency and order". *Language* 81: 96-137.

Felíu, E., Gil Laforga, I. and Torres, M. (eds.). 2006. *La Morfología a debate*. Jaén: Universidad de Jaén.

Ferrater, G. 1981. "La composició nominal". In Gabriel, F. (ed.). *Sobre el llenguatge*. Barcelona: Quaderns Crema. pp. 55-62.

Fodor, J. 1970. "Three Reasons for not Deriving *Kill* from *Cause to Die*". *Linguistic Inquiry* 1: 429-438.

Gaeta, L. 2006. "Lexical Integrity as a Constructional Strategy". *Lingue e Linguaggio* 1 [Special Issue on Lexical Integrity]: 67-82.

Gallardo, A. 1981. "Gramática de los nombres de colores". *Revista de lingüística teórica y aplicada* 19: 25-43.

Gallego, Á. J. 2007. *Phase Theory and Parametric Variation*. PhD dissertation, Universitat Autònoma de Barcelona.

Gallego, Á. J. 2008. "Phases and variation: Exploring the second factor of the language faculty". Ms. Universitat Autònoma de Barcelona.

García Lozano, F. 1978. "Los compuestos de sustantivo+adjetivo del tipo *pelirrojo*". In Varela, S. (ed.). *La formación de palabras*. 1993. Madrid: Taurus. pp. 205-214.

Gavarró, A. 1990a. "A reply to Scalise on 'The notion of "head" in morphology'". In Booij, G. and Marle, J. van (eds.). *Yearbook of Morphology* 3. Dordrecht: Foris. pp. 223-227.

Gavarró, A. 1990b. *Syntactic Theory and the Grammar of Catalan Compounds*. PhD dissertation, University of Edinburgh.

GDLC: *Gran Diccionari de la llengua catalana* http://www.enciclopedia.cat/

Giegerich, H. 1999. *Lexical Strata in English: Morphological causes, phonological effects*. Cambridge: Cambridge University Press.

Giegerich, H. 2004. "Compound or phrase? English noun-plus-noun constructions and the stress criterion". *English Language and Linguistics* 8: 1-24.

Giegerich, H. 2005. "Associative adjectives in English and the lexicon-syntax interface". *Journal of Linguistics* 41: 571-591.

Gil Laforga, I. 2003. "Estructura y significado de los adjetivos compuestos". *Interlingüística* 14: 439-446.

Gil Laforga, I. 2006. "Estructura Léxico-Sintáctica de los adjetivos compuestos de nombre y adjetivo". *Cuadernos de Lingüística del I. U. I. Ortega y Gasset* 13: 25-42.

Gleitman, L. R. and Gleitman, H. 1970. *Phrase and Paraphrase: Some Innovative Uses of Language*. New York: Norton.

Goldberg, A. E. 1995. *Constructions. A Construction Grammar Approach to Argument Structure*. Chicago & London: The University of Chicago Press.

González Calvo, J. M. 1976. "Sobre un tipo de construcción en la adjetivación de color". *Español Actual* 31: 56-58.

Gràcia, Ll. 1992. "-*ble* adjectives and middle constructions: a problem for inheritance". C*atalan Working Papers in Linguistics* 1992: 163-182.

Gràcia, Ll. 1995. *Morfologia Lèxica: l'herència de l'estructura argumental*. València: Servei de Publicacions de la Universitat de València.

Gràcia, Ll. 2002. "Formació de mots: composició". In Solà, J., Lloret, R.-M., Mascaró, J. and Pérez Saldanya, M. (eds.). *Gramàtica del català contemporani*. Barcelona: Empúries. pp: 777-829.

Gràcia, Ll. and Azkarate, M. 2000. "Prefixation and the head-complement parameter". In Dressler, W., Pfeiffer, O. E., Pöchtrager, M. and Rennison, J. R. (eds.). *Morphological Analysis in Comparison*. Amsterdam: John Benjamins. pp. 61-73.

Gràcia, Ll., Cabré, M. T., Varela, S. and Azkarate, M. 2000. *Configuración morfológica y estructura argumental: léxico y diccionario.* Zarautz (Guipúzcoa): Euskal Herriko Unibertsitateko Argitalpen Zerbitzua.

Gràcia, Ll. and Fullana, O. 1999. "On Catalan verbal compounds". *Probus* 11: 239-261.

Gràcia, Ll. and Fullana, O. 2000. "Catalan Verbal Compounds: Internal Order and Argument Interpretation". In Dressler, W., Pfeiffer, O. E., Pöchtrager, M. and Rennison, J. R. (eds.). *Morphological Analysis in Comparison.* Amsterdam: John Benjamins. pp. 75-87.

Grandi, N. 2009. "When Morphology 'Feeds' Syntax: Remarks on Noun > Adjective Conversion in Italian Appositive Compounds". In Montermini, F., Boyé, G. and Tseng, J. (eds.). *Selected Proceedings of the 6th Décembrettes: Morphology in Bordeaux.* Somerville, Mass.: Cascadilla Proceedings Project. pp. 112-124. http://www.lingref.com/cpp/decemb/6/

Grimshaw, J. 1986. "A Morphosyntactic Explanation for the Mirror Principle". *Linguistic Inquiry* 17: 745-749.

Grimshaw, J. 1990. *Argument Structure.* Cambridge, Mass.: MIT Press.

Grossmann, M. 1986. "Anàlisi dels compostos catalans del tipus *somiatruites*". In Camprubí, M. et al. (eds.). *Estudis de llengua i literatura catalanes* XII. Miscel·lània Antoni M. Badia i Margarit 4. Barcelona: Publicacions de l'Abadia de Montserrat. pp. 155-169.

Guerón, J. 1985. "Inalienable Possession, PRO-Inclusion and Lexical Chains". In Guerón, J., Obenauer, H.-G. and Pollock, J.-Y. (eds.). *Grammatical Representation.* Dordrecht: Foris Publications. pp. 43-86.

Guerón, J. 2003. "Inalienable possession and the interpretation of determiners". In Coene, M. and d'Hulst, Y. (eds.). *From NP to DP, vol.2: The expression of possession in noun phrases.* Amsterdam: John Benjamins. pp. 189-220.

Guevara, E. and Scalise, S. 2009. "Searching for Universals in Compounding". In Scalise, S., Magni, E., Vineis, E. and Bisetto, A. (eds.). *Universals of Language Today.* Amsterdam: Springer. pp. 101-128.

Hale, K. and Keyser, S. J. 1993. "On Argument Structure and the Lexical Expression of Syntactic Relations". In Hale, K. and Keyser, S. J. (eds.). *A view from Building 20: Essays in Linguistics in Honor of Sylvian Bromberger.* Cambridge, Mass.: MIT Press. pp. 53-109.

Hale, K and Keyser, S. J. 1998. "The Basic Elements of Argument Structure". In Harley, H. (ed.). *MIT Working Papers in Linguistics 32, Papers from the Upenn/MIT Roundtable on Argument Structure and Aspect.* Cambridge, Mass.: MIT Press. pp. 73-118.

Hale, K. and Keyser, S. J. 2000. "Aspect and the Syntax of Argument Structure". Ms. Cambridge, Mass.: MIT.

Hale, K. and Keyser, S. J. 2002. *Prolegomenon to a Theory of Argument Structure*. Cambridge Mass.: MIT Press.

Halle, M. 1973. "Prolegomena to a Theory of Word Formation". *Linguistic Inquiry* 4: 3-16.

Halle, M. 1997. "Distributed morphology: Impoverishment and fission". In Bruening, B., Kang, Y. and McGinnis, M. (eds.). *MITWPL 30: Papers at the Interface*. Cambridge, Mass.: MITWPL. pp. 425- 449.

Halle, M. and Marantz, A. 1993. "Distributed Morphology and the Pieces of Inflection". In Hale, K. and Keyser, S. J. (eds.). *A view from Building 20*: *Essays in Linguistics in Honor of Sylvian Bromberger*. Cambridge, Mass.: MIT Press. pp. 111-176.

Harley, H. 2002. "Why one head is better than two: Head-movement and compounding as consequences of Merge in Bare Phrase Structure". Talk given in the Arizona Linguistics Colloquium Series, University of Arizona, Arizona.

Harley, H. 2004. "Merge, conflation, and head movement: The First Sister Principle revisited". In Moulton, K. and Wolf, M. (eds.). *Proceedings of NELS* 34. University Massachusetts Amherst: GSLA. pp. 239-254. [cited by manuscript version]

Harley, H. 2006. *English Words*. Malden, MA et al.: Blackwell.

Harley, H. 2007. "The bipartite structure of verbs cross-linguistically (or: Why Mary can't *exhibit John her paintings*). Talk given at the 2007 ABRALIN Congress, Belo Horizonte, Brazil.

Harely, H. 2008a. "Bare Roots, Conflation and the Canonical Use Constraint". Talk given at the NORMS Workshop on Argument Structure, University of Lund, Sweden.

Harley, H. 2008b. "Compounding in Distributed Morphology". Ms. University of Arizona. [The final version can be found in: Lieber, R. and Štekauer, P. (eds.). 2009. *The Oxford Handbook of Compounding*. Oxford: Oxford University Press. pp. 129-144.]

Harley, H. and Noyer, R. 1998. "Licensing in the non-lexicalist lexicon: nominalizations, Vocabulary Items and the Encyclopedia". In Harley, H. (ed.). *MITWPL 32: Papers from the UPenn/MIT Roundtable on Argument Structure and Aspect*. Cambridge, Mass.: MITWPL. pp. 119-137.

Harley, H. and Noyer, R. 1999. "State-of-the-article: Distributed Morphology". *Glot International* 4: 3-9.

Harley, H. and Noyer, R. 2000. "Formal vs. Encyclopaedic Properties of Vocabulary. Evidence from Nominalizations". In Peeters, B. (ed.). *The Lexicon/Encyclopedia Interface*. Amsterdam: Elsevier Press. pp. 349-374.

Haspelmath, M. 2004. *Coordinating Constructions*. Amsterdam: John Benjamins.

Hatcher, A. G. 1960. "An Introduction to the Analysis of English Noun Compounds". *Word Journal of the International Linguistic Association* 16: 356-373.

Haugen, J. 2009. "Hyponymous objects and Late Insertion". *Lingua* 119: 242-262.

Hay, J. 2000. *Causes and Consequences of Word Structure*. PhD dissertation, Northwestern University.

Hay, J. 2002. "From Speech Perception to Morphology: Affix Ordering Revisited". *Language* 78: 527-555.

Hoeksema, J. 1988. "Head-types in morpho-syntax". In Booij, G. and van Marle, J. (eds.). *Yearbook of Morphology* 1988. Dordrecht: Foris. pp. 123-137.

Hoeksema, J. 1992. "The Head Parameter in Morphology and Syntax". In Gilbers, D. G. and Looyenga, S. (eds.). *Language and Cognition* 2. Groningen: De Passage. pp. 119-132.

Hoekstra, T. 1988. "Small Clause Results". *Lingua* 74: 101-139.

Hoekstra, T. and Mulder, J. 1990. "Unergatives as copular verbs". *The Linguistic Review* 7: 1-79.

Hoekstra, T. and van der Putten, F.. 1988. "Inheritance Phenomena". In Everaert, M., Evers, A., Huybregts, R. and Trommelen, M. (eds.). *Morphology and Modularity*. Dordrecht: Foris. pp. 163-186.

Holmes, P. and Hinchliffe I. 1994. *Swedish: A Comprehensive Grammar*. London: Routledge.

Horrocks, G. and Stavrou, M. 2007. "Grammaticalized aspect and spatio-temporal culmination". *Lingua* 117: 605-644.

http://www.askoxford.com/

Hualde, J. I. (ed.). 1998. Special issue *Italian Journal of Linguistics / Rivista di Linguistica: Metaphony and Vowel Harmony in Romance and Beyond* 10.

Huddleston, R. D. and Pullum, G. K. 2002. *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.

Hudson, R. A. 1987. "Zwicky on heads". *Journal of Linguistics* 23: 109-132.

Iacobini, C. and Masini, F. 2007. "The emergence of verb-particle constructions in Italian: locative and actional meanings". *Morphology* 16: 155-188.

Jackendoff, R. 1990. *Semantic Structures*. Cambridge, Mass.: MIT Press.

Jackendoff, R. 1992. "Babe Ruth Homered his Way into the Hearts of America". *Syntax and Cognition* 26: 155-178.

Jackendoff, R. 1997. *The Architecture of the Language Faculty*. Cambridge, Mass.: MIT Press.

Jackendoff, R. 2002. *Foundations of Language: Brain, Meaning, Grammar, Evolution*. Oxford: Oxford University Press.

Jackendoff, R. 2007. "Compounding in the Parallel Architecture and Conceptual Semantics". Ms. [The final version can be found in: Lieber, R. and Štekauer, P. (eds.). 2009. *The Oxford Handbook of Compounding*. Oxford: Oxford University Press. pp. 105-128.]

Jaeggli, O. 1986. "Three Issues in the Theory of Clitics: Case, Doubled NPs and Extraction". In Borer, H. (ed.). *The Syntax of Pronominal Clitics*. New York: Academic Press. pp. 15-42.

Julien, M. 2002. *Syntactic Heads and Word Formation*. Oxford: Oxford University Press.

Katamba, F. and Stonham, J. 2006. *Morphology*. London: McMillan. (2nd edition)

Kaufmann, I. and Wunderlich, D. 1998. "Cross-linguistic Patterns of Resultatives". Ms. University of Düsseldorf.

Kayne, R. S. 2009. "Antisymmetry and the Lexicon". *Linguistic Variation Yearbook 8*: 1-31.

Kiparsky, P. 1982. "Lexical phonology and morphology". In Yang, I. S. (ed.). *Linguistics in the Morning Calm*. Seoul: Hanshin. pp. 3-91.

Kiparsky, P. 1997. "Remarks on Denominal Verbs". In Alsina, A., Bresnan, J. and Sells, P. (eds.). *Complex predicates*. Stanford, CA: CSLI Publications. pp. 473-499.

Klingebiel, K. 1988. "New compounds from the old: an unexpected source of verb+noun compounds in Romance". *Berkeley Linguistic Society* 14: 88-99.

Kornfeld, L. 2003. "N+N Compounds as Formally Lexicalized Appositions in Spanish". In Booij, G., De Cesaris, J., Ralli, A. and Scalise, S. (eds). *Topics in Morphology: Selected Papers from the Third Mediterranean Morphology Meeting*. Barcelona: IULA. pp. 211-225.

Kornfeld, L. and Saab, A. L. 2003. "Morphology and Syntax: Prepositional Prefixes in Spanish". In Booij, G., De Cesaris, J., Ralli, A. and Scalise, S. (eds). *Topics in Morphology: Selected Papers from the Third Mediterranean Morphology Meeting*. Barcelona: IULA. pp. 227-240.

Kratzer, A. 2005. "Building Resultatives". In Maienborn, C. and Wöllstein, A. (eds.). *Event Arguments: Foundations and Applications*. Tübingen: Niemeyer. pp. 177-212.

Kubozono, H. 1990. "Phonological constraints on blending in English as a case for phonology-morphology interface". In Booij, G. and van Marle, J. (eds.). *Yearbook of Morphology* 1990. Dordrecht: Foris. pp. 1-20.

Ladd, D. R. 1984. "English compound stress". In Gibbon, D. and Richter, H. (eds.). *Intonation, accent and rhythm*. Berlin and New York: Walter de Gruyter. pp. 253-266.

Lang, M. F. 1992. *Formación de palabras en español*. Madrid: Cátedra.

Langacker, R. W. 1999. "The meaning of *of*". In Langacker, R. W. (ed.). *Grammar and Conceptualization*. Berlin: Mouton de Gruyter. pp. 73- 90.

Lardiere, D. 1998. "Parameter-Resetting in Morphology". In Beck, M.-L. (ed.). *Morphology and its interfaces in second language knowledge*. Amsterdam: John Benjamins. pp. 283-305.

Lees, R. B. 1960. *The Grammar of English nominalizations*. The Hague: Mouton.

Levi, J. 1978. *The Syntax and Semantics of Complex Nominals*. New York: Academic Press.

Levin, B. and Rapoport, T. R. 1988. "Lexical Subordination". *Chicago Linguistics Society* 24: 275-289.

Levin, B. and Rappaport, M. 1988. "Non-event –*er* Nominals: a Probe into Argument Structure". *Linguistics* 26: 1067-1083.

Levin, B. and Rappaport Hovav, M. 1995. Unaccusativity: At the Syntax-Lexical Semantics Interface. Cambridge, Mass.: MIT Press.

Levin, B. and Rappaport Hovav, M. 2005. *Argument Realization*. Cambridge/New York: Cambridge University Press.

Levinson, L. 2010. "Arguments for pseudo-resultative predicates". *Natural Language & Linguistic Theory* 28: 135-182.

Li, C. N. and Thompson, S. A. 1981. *Mandarin Chinese: a functional reference grammar*. Berkeley: University of California Press.

Liberman, M. and Prince, A. 1977. "On stress and linguistic rhythm". *Linguistic Inquiry* 8: 249-336.

Liberman, M. and Sproat, R. 1992. "The stress and structure of modified noun phrases in English". In Sag, I. A. and Szabolcsi, A. (eds.). *Lexical matters*. Stanford: Stanford University Press. pp. 131-181.

Liceras, J. M., Díaz, L. and Salomaa-Robertson, T. 2002. "The Compounding Parameter and the Word-Marker Hypothesis: Accounting for adult L2 Acquisition of Spanish N-N Compounds". In Pérez-Leroux, A. T. and Liceras, J. M. (eds.). *The Acquisition of Spanish Morphosyntax*. Dordrecht: Kluwer. pp. 209-237.

Lieber, R. 1981. *On the Organization of the Lexicon*. Bloomington: Indiana University Linguistics Club.

Lieber, R. 1983. "Argument linking and compounds in English". *Linguistic Inquiry* 14: 251-286.

Lieber, R. 1989. "On Percolation". In Booij, G. and van Marle, J. (eds.). *Yearbook of Morphology* 1989. Dordrecht: Kluwer. pp. 95-138.

Lieber, R. 1992. *Deconstructing Morphology*. Chicago/London: The University of Chicago Press.

Lieber, R. 2003. "Compound Interpretation: Lexical Semantics, Not Syntax". In Booij, G., De Cesaris, J., Ralli, A. and Scalise, S. (eds). *Topics in Morphology: Selected Papers from the Third Mediterranean Morphology Meeting*. Barcelona: IULA. pp. 241-253.

Lieber, R. 2004. *Morphology and Lexical Semantics*. Cambridge/New York: Cambridge University Press.

Lieber, R. 2006. "The category of roots and the roots of categories: what we learn from selection in derivation". *Morphology* 16: 247-272.

Lieber, R. 2008. "On the Lexical Semantics of English Compounds". Talk given at the CompoNet Congress on Compounding 2008, University of Bologna, Bologna.

Lieber, R. and Baayen, H. 1993. "Verbal Prefixes in Dutch: A Study in Lexical Conceptual Structure". In Booij, G. and van Marle, J. (eds.). *Yearbook of Morphology* 1993. Dordrecht: Kluwer. pp. 51-78.

Lieber, R. and Scalise, S. 2006. "The Lexical Integrity Hypothesis in a New Theoretical Universe". *Lingue e Linguaggio* 1 [Special Issue on Lexical Integrity]: 7-32.

Lieber, R. and Štekauer, P. 2009. "Introduction: status and definition of compounding". In Lieber, R. and Štekauer, P. (eds.). *The Oxford Handbook of Compounding*. Oxford: Oxford University Press. pp. 3-18.

Llinàs-Grau, M. 1997."Verb-complement patterns in early Catalan". In Hollebrandse, B. (ed.). *New Perspectives on Language Acquisition*. GLSA, Amherst: University of Massachusetts. pp. 15-27.

Llinàs-Grau, M. and Coll-Alfonso, M. 2001. "Telic verbs in early Catalan". *Probus* 13: 69-79.

Lorente, M. 2002. "Altres elements lèxics". In Solà, J., Lloret, R.-M., Mascaró, J. and Pérez Saldanya, M. (eds.). *Gramàtica del català contemporani*. Barcelona: Empúries. pp. 831-888.

Lyons, J. 1977. *Semantics*. Cambridge: Cambridge University Press.

Marantz, A. 1997a. "No Escape from Syntax: Don't Try Morphological Analysis in the Privacy of Your Own Lexicon". *UPenn Working Papers in Linguistics* 4: 201-225.

Marantz, A. 1997b. "Cat' as a phrasal idiom". Ms. MIT.

Marantz, A. 2001. "Words". Ms. MIT.

Marantz, A. 2003. "Subjects and Objects". Ms. NYU.

Marantz, A. 2007. "Phases and Words". Ms. NYU. [A published version can be found in: Choe, S. H. (ed.). *Phases in the theory of grammar*. Seoul: Dong In. pp. 191-220.]

Marchand, H. 1969. *The Categories and Types of Present-Day English Word-Formation.* München: CH Beck.

Marchis, M. 2009. "Compounding across languages". Talk given at Going Romance 2009, Université de Nice, Nice.

Martí, N. 2002. "El SN: els noms". In Solà, J., Lloret, R.-M., Mascaró, J. and Pérez Saldanya, M. (eds.). *Gramàtica del català contemporani*. Barcelona: Empúries. pp. 1281-1335.

Mascaró, J. 1986. *Morfologia*. Barcelona: Enciclopèdia Catalana.

Mascaró, J. 2002. "El sistema vocàlic. Reducció vocàlica". In Solà, J., Lloret, R.-M., Mascaró, J. and Pérez Saldanya, M. (eds.). *Gramàtica del català contemporani*. Barcelona: Empúries. pp. 89-123.

Masini, F. 2005. "Multi-word expressions between syntax and the lexicon: the case of Italian verb-particle constructions". *SKY Journal of Linguistics* 18: 145-173.

Mateu, J. 1999. "Universals of Semantic Construal for Lexical-Syntactic Relations". Talk given at the 1999 GLOW Workshop: Sources of Universals, University of Postdam, Postdam. Distributed as GGT-99-4 Research Report, Universitat

Autònoma de Barcelona. [A published version can be found in: Mateu. J. 2003. *Linguistic Research* 7: 45-80.]

Mateu, J. 2000. "Why Can't We Wipe the Slate Clean? A Lexical-Syntactic Approach to Resultative Constructions". *Catalan Working Papers in Linguistics* 8: 71-95.

Mateu, J. 2001. "Preverbs in Complex Denominal Verbs: Lexical Adjuncts or Core Predicates?". *Catalan Working Papers in Linguistics* 9: 37-51.

Mateu, J. 2002. *Argument Structure: Relational Construal at the Syntax-Semantics Interface*. PhD dissertation, Universitat Autònoma de Barcelona.

Mateu, J. 2005. "Impossible Primitives". In Werning, M., Machery, E. and Schurz, G. (eds.). *The Composition of Meaning and Content: Foundational Issues*. Frankfurt: Ontos. pp. 213-229.

Mateu, J. 2008. "On the l-syntax of directionality/resultativity: The case of Germanic preverbs". In Asbury, A., Dotlaĉil, J., Gehrke, B. and Nouwen, R. (eds.). *Syntax and Semantics of Spatial P*. Amsterdam: John Benjamins. pp. 221-250.

Mateu, J. 2009. "Events in non-verbal domains: a DM-based approach". Talk given at the Workshop on Events across categories, CCHS-CSIC, Madrid.

Mateu, J. 2010. "Two Types of Small Clause Results: Weak Resultatives Involve Incorporation, While Strong Resultatives Involve Conflation". Ms. UAB.

Mateu, J. and Amadas, L. 2001. "Syntactic Tools for Semantic Construal". Talk given at the 1st Conference on Tools in Linguistic Theory (TilT), Utrecht Institute of Linguistics OTS, Utrecht.

Mateu, J. and Rigau, G. 2002. "A Minimalist Account of Conflation Processes: Parametric Variation at the Lexicon-Syntax Interface". In Alexiadou, A. (ed.). *Theoretical Approaches to Universals*. Amsterdam: John Benjamins. pp. 211-236.

Mateu, J. and Rigau, G. 2009. "Romance Paths as Cognate Complements". In Masullo, P. J., O'Rourke, E. and Huang, C-H. (eds.). *Romance Linguistics: Structures, Interfaces, and Microparametric Variation*. Amsterdam/Philadelphia: John Benjamins. pp. 221-235.

Mateu, J. and Rigau, G. 2010. "Verb-particle constructions in Romance: a lexical-syntactic account". Ms. UAB. To appear in *Probus*.

Matthews, P. H. 1974/1991. *Morphology: An Introduction to the theory of word-structure*. Cambridge: Cambridge University Press. (1991: 2nd edition)

Matushansky, O. 2006. "Head-movement in linguistic theory". *Linguistic Inquiry* 37: 69-109.

McIntyre, A. 2004. "Event Paths, Conflation, Argument Structure and VP Shells". *Linguistics* 42: 523-571.

McIntyre, A. 2009. "Synthetic Compounds and Argument Structure: Messages from a *bandwagon-jumper-onner* and a *two-cents-worth-thrower-inner*". Talk given at the Roots Workshop, University of Stuttgart, Stuttgart.

Meibauer, J. 2007. "How marginal are phrasal compounds? Generalized insertion, expressivity, and I/Q-interaction". *Morphology* 17: 233-259.

Meyer, R. 1993. *Compound Comprehension in Isolation and in Context*. Tübingen: Niemeyer.

Mithun, M. 1984. "The Evolution of Noun Incorporation". *Language* 60: 847-894.

Mithun, M. 1986. "On the Nature of Noun Incorporation". *Language* 62: 32-37.

Moll, F. de B. 1952. *Gramática histórica catalana*. Madrid: Gredos.

Moll, F. de B. 1975. *Gramàtica catalana referida especialment a les Illes Balears*. Palma de Mallorca: Moll. (2nd edition)

MWOD: *Merriam-Webster Online Dictionary* http://www.merriam-webster.com/

Napoli, D. J. 1992. "Secondary resultative predicates in Italian". *Journal of Linguistics* 28: 53-90.

Neeleman, A. 1994. *Complex predicates*. PhD dissertation, Utrecht University.

Neeleman, A. and Schipper, J. 1992. "Verbal prefixation in Dutch: thematic evidence for conversion". In Booij, G. and van Marle, J. (eds.). *Yearbook of Morphology*. Dordrecht: Kluwer. pp. 57-92.

Neeleman, A. and van de Koot, H.. 2002a "Configurational Matrix". *Linguistic Inquiry* 33: 529-574.

Neeleman, A. and van de Koot, H. 2002b. "Bare Resultatives". *Journal of Comparative Germanic Linguistics* 6: 1-52.

Nespor, M. and Vogel, I. 1986. *Prosodic Phonology*. Dordrecht: Foris.

Nicholas, N. and Joseph, B. D. 2007. "Verbal Dvandvas in Modern Greek". Ms. To appear in *MIT Working Papers in Linguistics* (Papers from the MIT Workshop on Greek Syntax and Semantics (May 2007)).

Noyer, R. 1993. "Optimal Words: Towards a Declarative Theory of Word Formation". Ms. Princeton University.

Noyer, R. 1997. *Features, Positions and Affixes in Autonomous Morphological Structure*. New York: Garland.

Olsen, S. 2000a. "Compounding and Stress in English". *Linguistische Berichte* 181: 55-69.

Olsen, S. 2000b. "Composition". In Booij, G., Lehmann, Ch. and Mugdan, J. (eds.). *Morphology: An International Handbook on Inflection and Word-Formation*. Berlin: Walter de Gruyter. pp. 897-915.

Olsen, S. 2001. "Copulative compounds: a closer look at the interface between morphology and syntax." In Booij, G. and van Marle, J. (eds.). *Yearbook of Morphology* 2000. Dordrecht: Kluwer. pp. 279-320.

Olsen, S. 2004. "Coordination in morphology and syntax: the case of copulative compounds". In Meulen, A. ter and Abraham, W. (eds.). *The Composition of Meaning*. Amsterdam/Philadelphia: John Benjamins. pp. 17-38.

Oniga, R. 1988. *I composti nominali latini. Una morfologia generativa*. Bologna: Pàtron Editore.

Oniga, R. 1992. "Compounding in Latin". *Rivista di Linguistica* 4: 97-116.

Oshita, H. 1994. "Compounds: a view from suffixation and A-structure alteration". In Booij, G. and van Marle, J. (eds.). *Yearbook of Morphology*. Dordrecht: Kluwer. pp. 179-205.

Padrosa-Trias, S. 2005a. "Catalan Verbal Prefixation". In Daskalaki E., Katsos, N., Mavrogiorgos, M. and Reeve, M. (eds). *CamLing 2004: Proceedings of the University of Cambridge Second Postgraduate Conference in Language Research*. Cambridge: CILR. pp. 372-379.

Padrosa-Trias, S. 2005b. *Argument Structure and Morphology: the Case of en-Prefixation in English and Catalan*. MA dissertation, Universitat Autònoma de Barcelona.

Padrosa-Trias, S. 2006. "The derivation of en-prefixed verbs in Spanish". Ms. Universitat Autònoma de Barcelona.

Padrosa-Trias, S. 2007a. "Catalan Verbal Compounds and the Syntax-Morphology Competition". In Montermini, F., Boyé, G. and Hathout, N. (eds). *Selected Proceedings of the 5th Décembrettes: Morphology in Toulouse*. Somerville, Mass.: Cascadilla Proceedings Project. pp. 91-107.
http://www.lingref.com/cpp/decemb/5/index.html

Padrosa-Trias, S. 2007b. "Morpho-Syntactic Competition in Germanic and Romance Compounding". In Bisetto, A. and Barbieri, F. E. (eds.) *Proceedings of the XXXIII Incontro di Grammatica Generativa*. pp. 63-85.
http://amsacta.cib.unibo.it/archive/00002397/01/PROCEEDINGS_IGG33.pdf

Padrosa-Trias, S. 2008. "Some notes on the adjectival NA type of compound in Romance". Ms. Universitat Autònoma de Barcelona.

Padrosa-Trias, S. 2009a. "Exploring Compounding in Distributed Morphology". Poster presented at the 19th Colloquium on Generative Grammar, University of the Basque Country, Vitoria-Gasteiz.

Padrosa-Trias, S. 2009b. "Compounds in Distributed Morphology: Pros and Cons". Talk given in the LyCC Colloquium Series, CCHS-CSIC Seminar on Theoretical Linguistics, Madrid.

Padrosa-Trias, S. 2010. "Are there coordinate compounds". To appear in the *Mediterranean Morphology Meeting 7 Proceedings*, University of Cyprus.

Padrosa-Trias, S. (in press, a). "Argument Structure and Morphology: the Case of en-Prefixation Revisited". To appear in *ASJU-International Journal of Basque Linguistics and Philosophy*.

Padrosa-Trias, S. (in press, b). "*Entre*-prefixed Verbs in Catalan: a Lexical Semantic Account". *Interlingüística* 18.

Padrosa-Trias, S. and Markova, A. 2009. "Some Remarks on Prefixation: Evidence from Bulgarian, Catalan and English". *Interlingüística* 19: 200-213.

Pérez Saldanya, M., Sifre, M. and Todolí, J. 2004. *Morfologia catalana*. Barcelona: UOC.

Pesetsky, D. 1995. *Zero syntax: experiencers and cascades*. Cambridge, Mass.: MIT Press.

Picallo, C. 1991. "Nominals and Nominalizations in Catalan". *Probus* 33: 279-316.

Picallo, C. 2002. "L'adjectiu i el sintagma adjectival". In Solà, J., Lloret, R.-M., Mascaró, J. and Pérez Saldanya, M. (eds.). *Gramàtica del català contemporani*. Barcelona: Empúries. pp. 1641-1688.

Picallo, C. 2008. "Gender and number in Romance". *Lingue e Linguaggio* 7: 47-66.

Piera, C. 1995. "On Compounding in English and Spanish". In Campos, H. and Kempchinsky, P. (eds.). *Evolution and Revolution in Linguistic Theory*. Washington, D.C.: Georgetown University Press. pp. 302-315.

Piera, C. and Varela, S. 1999. "Relaciones entre morfología y sintaxis". In Bosque, I. and Demonte, V. (eds.). *Gramática Descriptiva de la Lengua Española*. Madrid: Espasa Calpe. pp. 4367-4422.

Pinker, S. 1994. *The Language Instinct*. London: Penguin Books.

Pinker, S. 1999. *Words and Rules: the ingredients of language*. New York: HarperCollins.

Pinkster, H. 1990. *Latin syntax and semantics*. London: Routledge.

Plag, I. 2003. *Word-Formation in English*. Cambridge: Cambridge University Press.

Prieto, P. 2003. "Correlats acústics de l'accent secundari en català". *Estudios de Fonética Experimental* 12: 107-142.

Prieto, P. 2004. *Fonètica i fonologia. Els sons del català*. Barcelona: UOC.

Pullum, G. K. and Scholz, B. C. 2002. "Empirical assessment of stimulus poverty arguments." *The Linguistic Review* 19: 9-50.

Pustejovsky, J. 1995. *The Generative Lexicon*. Cambridge, Mass.: MIT Press.

Pylkkänen, L. 2002. *Introducing Arguments*. PhD dissertation, MIT.

Radford, A. 1988. *Transformational Grammar: a first course.* Cambridge: Cambridge University Press.

Rainer, F. 1993. *Spanische Wortbildungslehre*. Tubinga: Niemeyer.

Rainer, F. and Varela, S. 1992. "Compounding in Spanish". *Rivista di linguistica* 4: 117-142.

Ralli, A. and Stavrou, M. 1997. "Morphology-syntax interface: A-N compounds vs. A-N constructs in Modern Greek". In Booij, G. and van Marle, J. (eds). *Yearbook of Morphology* 1998. Dordrecht: Kluwer. pp. 243-264.

Rappaport Hovav, M. and Levin, B. 1998. "Building Verb Meaning". In Butt, M and Geuder, W. (eds.). *The Projection of Arguments*. Stanford: CSLI Publications. pp. 97-134.

Recasens, D. 1993. *Fonètica i fonologia catalanes*. Barcelona: Enciclopèdia Catalana.

Reinhart, T. 2000. "The Theta System: Syntactic Realization of Verbal Concepts". In *OTS Working Papers in Linguistics*. To appear (extended) in the LI Monographs Series, MIT Press.

Reinhart, T. 2001. "A Synopsis of the 'Theta system'". Ms. Utrecht University. [An extended version can be found in: Reinhart, T. 2002. "The Theta System – an Overview". *Theoretical Linguistics* 28: 229-290.]

Rice, S. and Prideaux, G. 1991. "Event-packing: the case of object incorporation in English". *Berkeley Linguistic Society* 17: 283-298.

Rigau, G. 2002. "Els complements adjunts". In Solà, J., Lloret, R.-M., Mascaró, J. and Pérez Saldanya, M. (eds.). *Gramàtica del català contemporani*. Barcelona: Empúries. pp. 2045-2110.

Rizzi, L. 1986. "Null Objects in Italian and the theory of pro". *Linguistic Inquiry* 17: 501-557.

Roeper, T. and Siegel, M. 1978. "A lexical transformation for verbal compounds". *Lingusitic Inquiry* 9: 199-260.

Roeper, T. and Snyder, W. 2005. "Language learnability and the forms of recursion". In Di Sciullo, A. M. (ed.). *UG and External Systems: Language, Brain and Computation*. Amsterdam: John Benjamins. pp. 155-169.

Roeper, T., Snyder, W. and Hiramatsu, K. 2002. "Learnability in a Minimalist Framework: Root Compounds, Merger, and the Syntax-Morphology Interface". In Lasser, I. (ed.). *The Process of Language Acquisition*. Frankfurt: Peter Lang Verlag.

Romero, J. 1997. *Construcciones de doble objeto y gramática universal*. PhD dissertation, Universidad Autónoma de Madrid.

Rumelhart, D. E. and McClelland, J. L. 1986. "On Learning the Past Tenses of English Verbs". In McClelland, J. L., Rumelhart, D. E. and the PDP Research Group (eds.). *Parallel Distributed Processing*, vol. 2. Cambridge, Mass./London: MIT Press. pp. 216-271.

Sánchez López, C. 2003. "La relación de posesión inalienable en los compuestos". In Girón Alconchel, J. L., Herrero, J., Iglesias, S. and Narbona, A. (eds.). *Estudios ofrecidos al profesor José Jesús de Bustos Tovar*, vol. 1. Madrid: Editorial Complutense. pp.157-169.

Sato, Y. 2007. "Phrasal Compounding and Lexicalism: A Multiple Spell-Out Account". Ms. University of Arizona.

Scalise, S. 1984. *Generative Morphology*. Dordrecht: Foris.

Scalise, S. 1988. "The notion of 'head' in morphology". In Booij, G. and van Marle, J. (eds.). *Yearbook of Morphology* 1. Dordrecht: Foris. pp. 229-245.

Scalise, S. 1992. "Compounding in Italian". *Rivista de Linguistica* 4: 175-199.

Scalise, S. 2008. "Why compounding?" Talk given at the CompoNet Congress on Compounding 2008, University of Bologna, Bologna.

Scalise, S. and Bisetto, A. 2009. "The classification of compounds". In Lieber, R. and Štekauer, P. (eds.). *The Oxford Handbook of Compounding*. Oxford: Oxford University Press. pp. 34-53.

Scalise, S., Bisetto, A. and Guevara, E. 2005. "Selection in Compounding and Derivation". In Dressler, W. U., Kastovsky, D., Pfeiffer, O. E. and Rainer, F. (eds.). *Morphology and its Demarcation*. Amsterdam/Philadelphia: John Benjamins. pp. 133-150.

Scalise, S. and Guevara, E. 2006. "Exocentric compounding in a typological framework". *Lingue e Linguaggio* 2: 185-206

Scholz, C. 2009. "El Pinchadiscos: Romance Verb Noun Compounds". Talk given at Going Romance 2009, Université de Nice, Nice.

Selkirk, E. 1982. *The Syntax of Words*. Cambridge, Mass.: MIT Press.

Selkirk, E. 1984. *Phonology and Syntax*. Cambridge, Mass.: MIT Press.

Selkirk, E. 1986. "On derived domains in sentence phonology". *Phonology Yearbook* 3: 371-405.

Shim, J. Y. and Dikken, M. den 2007. "The Tense of Resultatives: the Case of Korean". Ms. CUNY Graduate Center. [A published version can be found in *NELS* 38: 337-350.]

Siegel, D. 1978. "The Adjacency Constraint and the Theory of Morphology". In Stein, M. (ed.). *Proceedings of the Eighth Annual Meeting of the North Eastern Linguistic Society* 8. Amherst, Mass.: University of Massachusetts. pp: 189-197.

Siegel, D. 1979. *Topics in English Morphology*. New York/London: Garland.

Snyder, W. 1995. *Language Acquisition and Language Variation: The Role of Morphology*. PhD dissertation, MIT.

Snyder, W. 1996. "The Acquisitional Role of the Syntax-Morphology Interface: Morphological Compounds and Syntactic Complex Predicates". In Stringfellow, A., Cahana-Amitay, D., Hughes, E. and Zukowski, A. (eds.). *Proceedings of the 20th Annual Boston University Conference on Language Development*, vol.2. Somerville, Mass.: Cascadilla Press. pp. 728-735.

Snyder, W. 2001. "On the Nature of Syntactic Variation: Evidence from Complex Predicates and Complex Word-Formation". *Language* 77: 324-342.

Snyder, W. 2002. "Parameters: The View from Child Language". In Otsu, Y. (ed.). *Proceedings of the Third Tokyo Conference on Psycholinguistics*. Tokyo: Hituzi Shobo. pp. 27-44.

Snyder, W. 2005. "Motion Predicates and the Compounding Parameter: A New Approach". Talk given in the Linguistics Colloquium Series, University of Maryland, College Park.

Snyder, W., Felber, S., Kang, B. and Lillo-Martin, D. 2001. "Path phrases and compounds in the acquisition of English". Talk given at the 26th Boston University Conference on Language Development, Boston.

Son, M. 2007. "Directionality and Resultativity: the Cross-linguistic Correlation Revisited". *Tromsø Working Papers on Language and Linguistics: Nordlyd 34.2* (Special issue on Space, Motion, and Result): 126-164.

Spencer, A. 1988. "Bracketing Paradoxes and the English Lexicon". *Language* 64: 663-682.

Spencer, A. 1991. *Morphological Theory*. Oxford: Blackwell.

Spencer, A. 2000. "Morphology and Syntax". In Booij, G., Lehmann, Ch. and Mugdan, J. (eds.). *Morphology: An International Handbook on Inflection and Word-Formation*. Berlin: de Gruyter. pp. 312-335.

Spencer, A. 2003a. "In Defence of Morphology". *Lingue e Linguaggio* 2: 231-243.

Spencer, A. 2003b. "Does English Have Productive Compounding?". In Booij, G., De Cesaris, J., Ralli, A. and Scalise, S. (eds.). *Topics in Morphology: Selected Papers from the Third Mediterranean Morphology Meeting*. Barcelona: IULA. pp. 329-341.

Spencer, A. and Zaretskaya, M. 1998. "Verb prefixation in Russian as lexical subordination". *Linguistics* 36: 1-39.

Sproat, R. 1985. *On Deriving the Lexicon*. PhD dissertation, MIT.

Sproat, R. 1988. "Bracketing Paradoxes, Cliticization and Other Topics: The Mapping between Syntactic and Phonological Structure". In Everaert, M., Evers, A., Huybregts, R. and Trommelen, M. (eds.). *Morphology and Modularity*. Dordrecht: Foris. pp. 339-360.

Stiebels, B. 1998. "Complex Denominal Verbs in German and the Morphology-Semantics Interface". In Booij, G. and van Marle, J. (eds.), *Yearbook of Morphology 1997*. Dordrecht: Kluwer. pp. 265-302.

Svenonius, P. 1996. "The verb-particle alternation in the Scandinavian languages". Ms. University of Tromsø.

Talmy, L. 1985. "Lexicalization patterns: Semantic structure in lexical forms". In Shopen, T. (ed.). *Language Typology and Syntactic Description I: Clause Structure*. Cambridge: Cambridge University Press. pp. 57-149.

Talmy, L. 1991. "Paths to Realization: A Typology of Event Integration". *Buffalo Papers in Linguistics* 91: 147-187.

Talmy, L. 2000. *Toward a Cognitive Semantics, vol. 2*. Cambridge, Mass.: MIT Press.

Taylor, J. R. 1996. *Possessives in English: an exploration in cognitive grammar*. Oxford: The Clarendon Press.

Todolí, J. 2002. "Els pronoms". In Solà, J., Lloret, R.-M., Mascaró, J. and Pérez Saldanya, M. (eds.). *Gramàtica del català contemporani*. Barcelona: Empúries. pp. 1337-1433.

Tomioka, N. 2004. "The Lexicalization Patterns of Verbs and V-V Compounds". In Beermann, D. and Hellan, L. (eds.). *Proceedings of TROSS 03: Workshop on Multi-Verb-Constructions*. Trondheim. [cited by manuscript version]

Torner, S. 2005. "On the morphological nature of Spanish adverbs ending in –mente". *Probus* 17: 115-144.

Tubau, S. 2004. "Early Catalan OV Sequences: Empirical Evidence for the Poverty of Stimulus Argument". *UCL Working Papers in Linguistics* 16: 149-163.

Turon, L. 1999. *La construcció del significat en els mots complexos: semàntica dels derivats amb formes prefixades preposicionals*. MA dissertation, IULA, Universitat Pompeu Fabra.

Uriagereka, J. 1988. *On government*. PhD dissertation, University of Connecticut.

Val Álvaro, J. F. 1999. "La composición". In Bosque, I. and Demonte, V. (eds.). *Gramática Descriptiva de la Lengua Española*. Madrid: Espasa Calpe. pp. 4757-4841.

Val Álvaro, J. F. 2006. "La morfología como componente de la gramática". In Felíu, E., Gil, I. and Torres, M. (eds.). *La Morfología a debate*. Jaén: Universidad de Jaén. pp. 33-61.

Varela, S. 1989. "Spanish endocentric compounds and the 'atom condition'". In Kirschner, C. and DeCesaris, J. (eds.). *Studies in Romance Linguistics*. Amsterdam: John Benjamins. pp. 397-441.

Varela, S. 1990. "Composición nominal y estructura temática". *Revista Española de Lingüística* 20: 55-81.

Varela, S. (ed.). 1993. *La formación de palabras*. Madrid: Taurus.

Varela, S. 1999. "Sobre las relaciones de la morfología con la sintaxis". *Revista Española de Lingüística* 29: 257-281.

Varela, S. 2005. *Morfología léxica: la formación de palabras*. Madrid: Gredos.

Varela, S. and Martín, J. 1999. "La prefijación". In Bosque, I. and Demonte, V. (eds.). *Gramática Descriptiva de la Lengua Española*. Madrid: Espasa Calpe. pp. 4993-5040.

Vendler, Z. 1967. *Linguistics in Philosophy*. Ithaca, NY: Cornell University Press.

Vergnaud, J-R. and Zubizarreta, M. L. 1992. "The Definite Determiner and the Inalienable Constructions in French and in English". *Linguistic Inquiry* 23: 595-652.

Villalva, A. 1992. "Compounding in Portuguese". *Rivista di Linguistica* 4: 201-219.

Voyles, J. B. 1974. *West Germanic. Inflection, Derivation and Compounding*. The Hague: Mouton.

Wälchli, B. 2005. *Co-compounds and Natural Coordination*. Oxford: Oxford University Press.

Walinska de Hackbeil, H. 1985. "En-prefixation and the Syntactic Domain of Zero Derivation". *Proceedings of the 11th Annual Meeting of the Berkeley Linguistic Society*: 337-357.

Washio, R. 1997. "Resultatives, compositionality and language variation". *Journal of East Asian Linguistics* 6: 1-49.

Wasow, T. 1977. "Transformations and the Lexicon". In Culicover, P., Wasow, T. and Bresnan, J. (eds.). *Formal Syntax*. New York: Academic Press. pp. 327-360.

Wheeler, M. 1977. "Problemes lèxics per al generativista: els mots catalans a doble arrel". In Colon, G. (ed.). *Actes del Quart Col·loqui Internacional de Llengua i Literatura Catalanes*. Montserrat: Publicacions de l'Abadia de Montserrat. pp. 235-250.

Whelpton, M. 2007. "Building Resultatives in Icelandic". Ms. University of Iceland.

Wiese, R. 1996. "Phrasal Compounds and the Theory of Word Syntax". *Linguistic Inquiry* 27: 183-193.

Williams, E. 1981a. "On the Notions 'Lexically Related' and 'Head of a Word'". *Linguistic Inquiry* 12: 245-274.

Williams, E. 1981b. "Argument Structure and Morphology". *The Linguistic Review* 1: 81-114.

Williams, E. 1994. "Remarks on lexical knowledge". *Lingua* 92: 7-34.

Williams, E. 2007. "Dumping Lexicalism". In Ramchand, G. and Reiss, C. (eds.). *The Oxford Handbook of Linguistic Interfaces*. Oxford: Oxford University Press. pp. 353-381.

Williams, E. 2008. "COMBINE: Reconciling lexicalism and morphosyntax". Talk given at the Workshop Words don't come easy, University of Verona, Verona.

Wisniewski, E. 1996. "Construal and Similarity in Conceptual Combination". *Journal of Memory and Language* 35: 434-453.

Yoneyama, M. 1986. "Motion Verbs in Conceptual Semantics". *Bulletin of the Faculty of Humanities* 22: 1-15.

Zubizarreta, M. L. 1985. "The relation between Morphology and Morphosyntax: The case of Romance causatives". *Linguistic Inquiry* 16: 247-289.

Zubizarreta, M. L. and Oh, E. 2007. *On the Syntactic Composition of Manner and Motion*. Cambridge, Mass.: MIT Press.

Zwanenburg, W. 1992a. "Morphological heads, French 'compounding' and Germanic 'prefixation'". In Laeufer, C. and Morgan, T. A. (eds.). Theoretical *Analyses in Romance Linguistics*. Amsterdam/Philadephia: John Benjamins. pp. 167-179.

Zwanenburg, W. 1992b. "Compounding in French". *Rivista de Linguistica* 4: 221-240.

Zwicky, A. M. 1985. "Heads". *Journal of Linguistics* 21: 1-29.