

The classification of compounds
Antonietta Bisetto* and Sergio Scalise*

0. Introduction

Compounds are generally acknowledged to be formed by at least two words¹ (e.g. Eng. *taxi-driver*, *postman*, It. *divano-letto* ‘sofa-bed’, *camposanto* ‘cemetery’, Fr. *tire bouchon* ‘cork-screw’, *timbre-poste* ‘stamp’, etc.). Though this definition does not always hold true², it is nonetheless descriptively valid for the core of compounding processes.

Since we have reasons to believe that the classifications of compounds provided so far are not adequate, in this paper we would like to propose a new classification which is at the same time based on consistent criteria and not just valid for one language or for a family of languages, but hopefully universally valid³.

The paper is organized as follows: in section 1 we will discuss the main problems with current and previous classifications of compounds; in section 2 we will present our proposed classification and in section 3 we discuss some residual problems and draw some provisional conclusions.

* University of Bologna

We would like to thank Geert Booij, Emiliano Guevara, Rochelle Lieber and Nigel Vincent for reading a previous version of this paper and giving us valuable suggestions. Usual disclaimers apply. This paper is part of a larger research on compounding made possible also thanks to funds of the University of Bologna.

¹ Cf. among others Allen (1978), Selkirk (1982), Scalise (1992), ten Hacken (1994).

² Leaving aside recursive composition, there are at least two cases where this definition is not true: a) in the so called neoclassical compounds and b) in phrasal compounds. The former are constructions such as *anthropology* and *insecticide* where one or both constituents are elements of Greek or Latin origin (*anhtropo-* and *-logy* in the first example and *-cide* in the second one). The latter are those compounds found in Germanic languages whose first constituent is a phrasal construction as in [*God is dead*] *theology* (cf. Lieber 1992).

³ For this article, we will use data from the Morbo/Comp project, a data base of compounds under development at the Department of Foreign Languages in Bologna. We have tested our classification of compounds on the languages analyzed so far within the Morbo/Comp project (Italian, English, Dutch, Spanish, Chinese, French, German, Bulgarian, Norwegian, Russian).

1. Problems in compound classifications

Classifications of compounds found in the current literature present a range of different problems, namely: terminological problems (1.1.), problems of neglected lexical categories (1.2.), and problems of consistency of the criteria used (1.3., 1.3.1.).

1.1. Terminological problems

Terminological problems can be said to fall into two distinct types, basically caused by a) meaning shifts of the definitory terms in the course of time and b) the language specific nature of many terms.

To illustrate the first type of problems, one example will suffice. The term *bahuvrihi*, originally used to designate a possessive exocentric compound ((*one who has*) much rice) with time ended up having the only generic meaning of «exocentric». Or, as Bauer (2001:700) puts it, this term ended up applying «to any compound which is not a hyponym of its own head element»⁴.

As for the second type of problems, the Anglophone tradition in particular has taken into account almost exclusively two types of compounds: root (or primary) compounds and synthetic (or secondary) compounds. The notion of root compound, however, cannot be applied to all languages; Romance languages, for instance, do not have compounds based on roots because Romance lexemes, besides the root, can contain another element with grammatical information. In an Italian compound like *capostazione* ‘station master’, for example, the two words *capo* and *stazione* are formed by the two roots *cap-* and *stazion-* plus a grammatical morpheme, *-o* and *-e* respectively.

Also the label «synthetic compound» used for constructions such as *taxi driver*, is not universally applicable because, still in Romance languages, this type of compound does not seem to exist. *Root compound* and *synthetic compound* are thus language specific terms. This is probably the reason why these terms, while largely utilized during the '70/'80 (cf., among others, Allen 1978 and Selkirk 1982) have thereafter been abandoned in the morphological literature

⁴ Cf. for example Booij (2005: 80) according to whom *bahuvrihi* compounds «are sometimes considered to form a subset of the exocentric compounds».

Also problematic are labels such as «phrasal» and «neoclassical». In fact, *phrasal* refers to the (syntactic) nature of the non-head, while *neoclassical* refers to the (Greek or Latinate) nature of the head or of both constituents.

1.2. Neglected categories

Morphological research (especially on English compounds) has favoured some categories of compounds, in particular those formed by two Nouns or by an Adjective and a Noun (both having categorial output Noun). Other lexical categories (either as input or output) have often been neglected. This is the case a) of compound adjectives (e.g. the A+A (*bitter-sweet*), A+N (It. *giallo limone* 'lemon yellow'), V+A (Du. *druipnat* 'drip-wet') and N+A (*girl crazy*) types), b) of compounds containing i) adverbs (It. *sottosopra* 'upside down'), ii) prepositions (Fr. *sans papier* lit. 'without documents'), iii) pronouns (*self-determination*), iv) particles (*make-up*) or v) verbs (It. *portacenere* 'ash tray').

A quick look at the data analyzed in the framework of the Morbo/Comp project reveals that besides the structures [N+N] and [A+N], the world's languages can have many other compounding structures. Besides the ones just mentioned, compound structures such as [V+V]_V, [V+V]_N, [A+V]_V, [N+V]_N, [N+V]_V, [V+N]_N, [V+N]_V, [Pro+N]_N, [Pro+V]_V, [P+A]_A, [Adv+V]_V etc., can be found whose properties have been only sporadically studied. Among them, there are highly productive structures, as is the case for Italian V+N compounds.

1.3. Classificatory criteria

The most salient problem in the classification of compounds has to do with the heterogeneous nature of the criteria adopted. It is to this problem that we will devote our attention in the following section.

To illustrate the point, let us take into account the notions of endocentricity/exocentricity and the notion of coordination. The first two notions define compounds on the basis of the presence vs. absence of a head constituent; the notion of coordination is based on the grammatical relation holding between the two constituents of the compound. However, if we do not allow any kind of intersection of the different notions involved, we would be forced to say, for example, that coordination has no relation whatsoever with endocentricity/exocentricity. But, as we will see below,

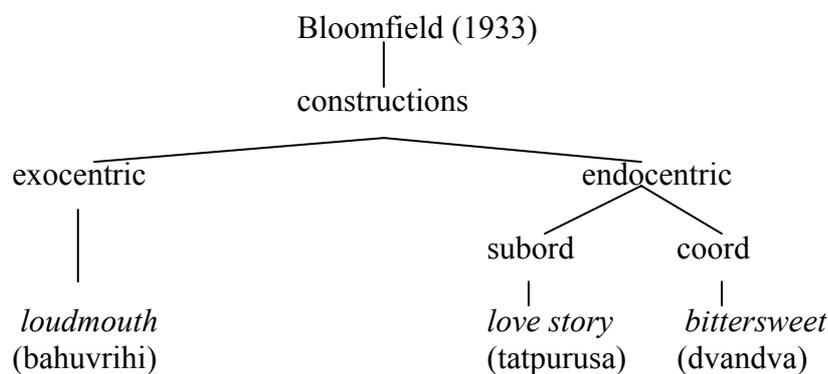
intersection of defining criteria is fundamental in order to obtain a descriptively adequate classification.

1.3.1. Previous and current classifications

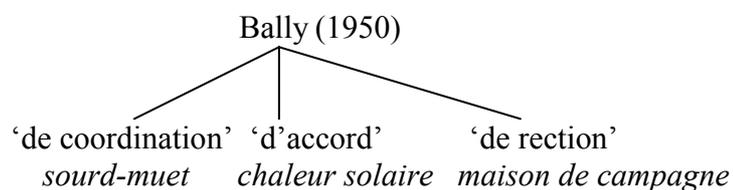
In order to illustrate the problems just mentioned, we will take into account an indicative if not exhaustive survey of some different proposals (that we illustrate in (1) making use of trees)⁵:

1)

a)

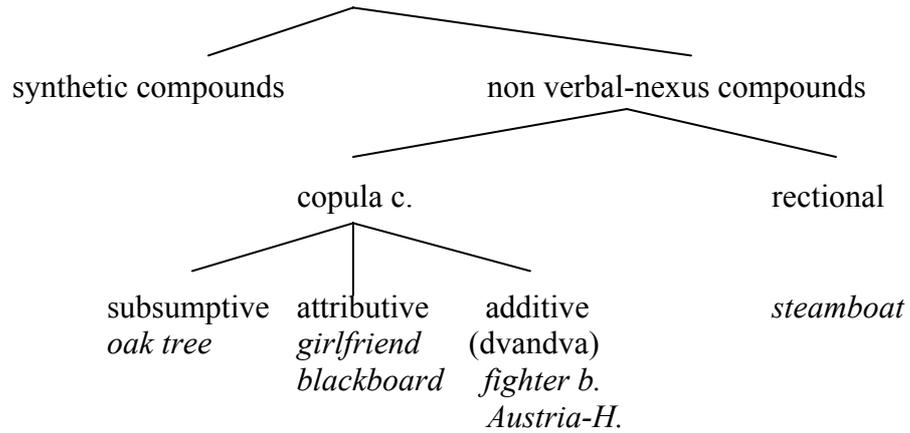


b)

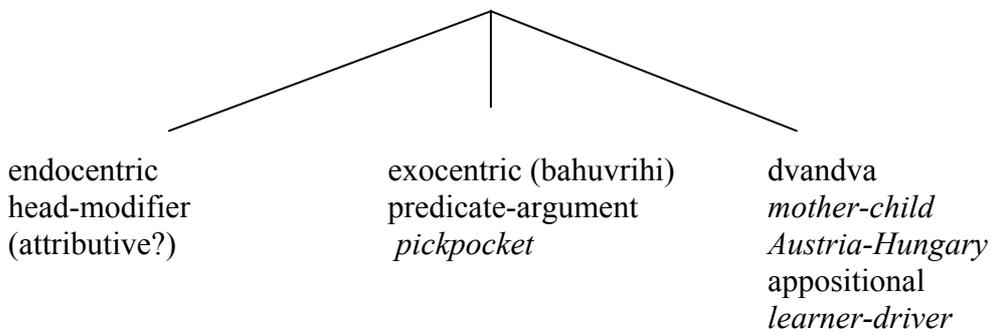


⁵ The representations given in (1a-i) are not taken in this form from the quoted works, but are the result of our interpretation, which in some cases is undoubtedly partial and schematic. In general, quoted works do not intend to propose a real classification of compounds but they represent, however, a state of the art of our knowledge on the topic.

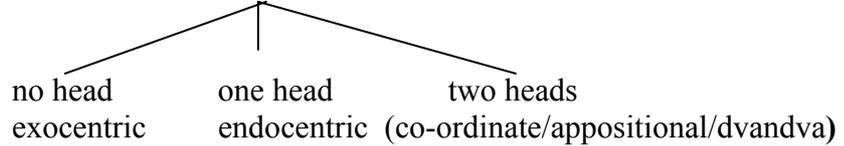
c) Marchand (1969) (only endocentric compounds)



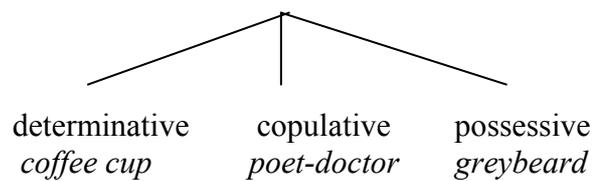
d) Spencer (1991)



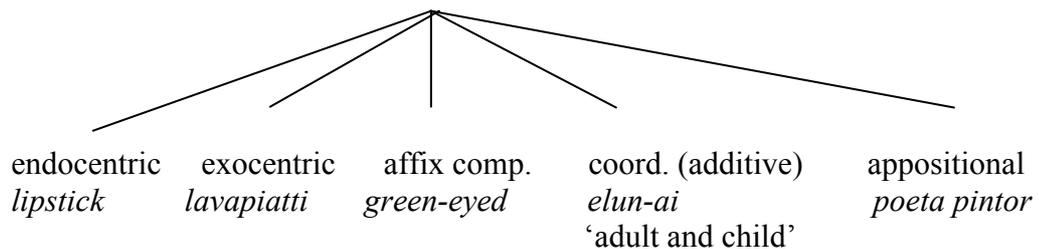
e) Fabb (1998)

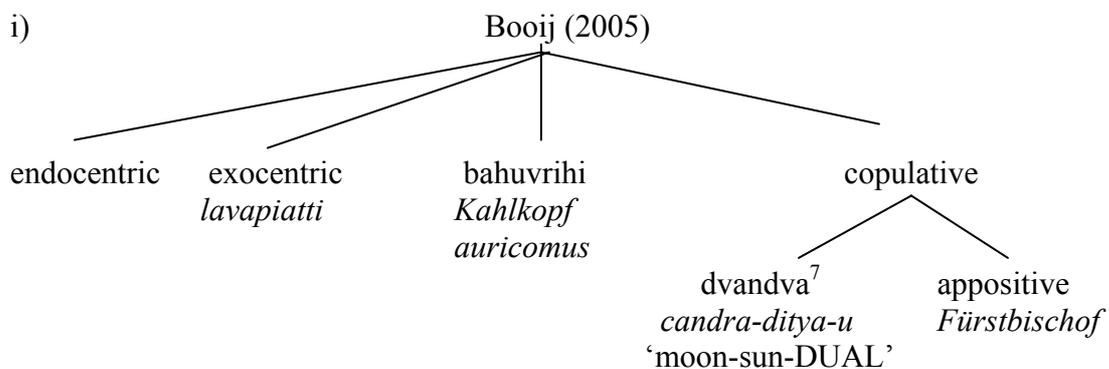
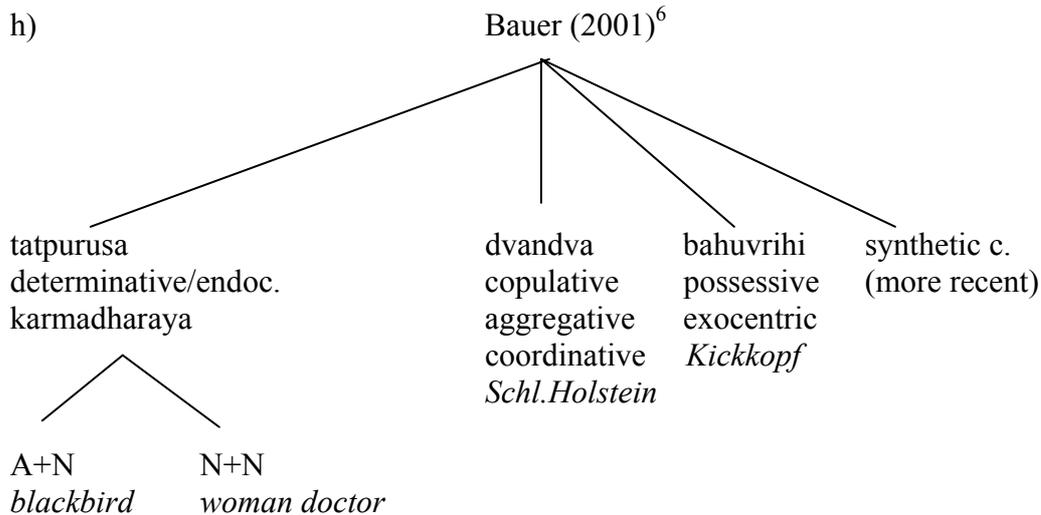


f) Olsen (2001)



g) Haspelmath (2002)





Let's first examine a set of proposals that seem *grosso modo* equivalent to each other, namely those of Spencer, Haspelmath, Booij and Bauer.

As can be easily observed, all four classifications consider the presence/absence of a head constituent as a criterion of the same level as, for instance, copulative (Booij and Bauer), dvandva (Spencer), coordinative and appositional (Haspelmath). In other terms, these scholars seem to set apart endocentricity and exocentricity, not allowing these two notions to extend across classes. Spencer, for instance, proposes that compounds are classed into 3 groups: a) endocentric head–modifier constructions – maybe also containing attributive compounds; b) exocentric (bahuvrihi) predicate-

⁶ Bauer's paper is a typical example of what we have said in footnote 5. Bauer in fact acknowledges that current classifications are problematic (for instance, in his view, a compound such as *woman doctor* should be considered as coordinative and not as *karmadharaya*). On the other hand, Bauer's paper has typological rather than classificatory aims.

⁷ Booij points out that «the copulative/appositive compounds [such as *Fürstbischhof*] are different from dvandva compounds because their number is singular».

argument formations and c) dvandva compounds – group perhaps including appositional constructions. But, separating endocentric and exocentric compounds from dvandvas has the undesired consequence that the latter seem to be unanalyzable on the basis of the presence or absence of the head. But dvandvas and appositionals are different just in that the former are exocentric and the latter endocentric. Not only head-modifier compounds, in fact, are endocentric and not only predicate-argument ones are exocentric.

Also Haspelmath's groups (endocentric, exocentric, coordinate, appositive and «affixed compounds») seem to obscure the fact that both affixed and coordinate (additive) compounds of the type *adult-child* are exocentric while appositional ones are endocentric.

In Booij's arrangement, though dvandva compounds of the Sanskrit type (with dual or plural inflection) are correctly separated from copulatives (which have singular number) the separation of endocentric and exocentric compounds from bahuvrihis and copulatives causes redundancy in that dvandva copulative compounds are exocentric while appositive copulative compounds are endocentric.

Similar observations can be made about the classification extrapolated from Bauer's work. Also in Bauer's taxonomy the notion of head does not apply with the due extension to all types of compounds.

Besides the problem we have just seen, which derives from the layering of notions like endocentricity and exocentricity on a par with other notions like dvandva and synthetic, a problem of inconsistency arises from these classifications.

Consider, for example, Haspelmath's proposal. His classification uses different classifying criteria: a) presence/absence of a head (giving rise to the distinction between endocentric and exocentric compounds), b) formal structure of compounds (introducing a class of «affixed compounds») and c) syntactic-semantic relation between constituents (determining the class of appositional compounds). As a consequence, it is not easy to understand whether or not criterion (a), i.e. presence/absence of a head, can be applied to compounds classed on the basis of criterion (b), i.e. formal structure, or whether there is any possible relation between affixed compounds and appositional ones.

Different observations can be made about Olsen's and Fabb's proposals. Fabb's classification, though consistent in the sense that it makes use of a single criterion –

number of heads – is too restrictive if confronted with the variety of attested compounds.

Olsen's classification has the advantage of using the notion of determinative as opposed to the notion of coordinative compounds. However, besides these two classes, Olsen introduces the class of (exocentric) possessive compounds, which is clearly a class based on a different criterion. An undesired consequence of this mixture of criteria is that it is not clear whether or not the notion endocentric/exocentric does apply to determinative and copulative compounds. Furthermore, while in determinative and copulative compounds the relevant relation is the one between the two constituents, in possessive compounds the «possessive» relation is the one between the whole compound and the absent head.

In passing, it can be noted that a label is not always used to indicate the same type of compounds (though this is normal in any scientific taxonomy). For example, what Olsen calls copulative compound (i.e. *poet-doctor*) is called appositional by Haspelmath and Spencer (i.e. *poeta-pintor*, *learner-driver*).

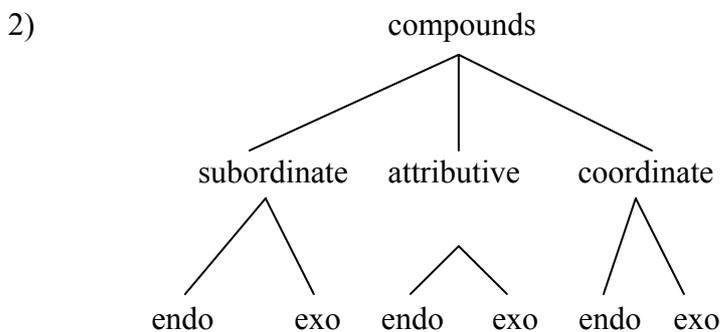
Let us now have a look at some previous «traditional» classifications. Bloomfield used consistently the notions of «subordinate» and «coordinate»; however, the structure of his classification misses the fact that both subordinate and coordinate compounds can be exocentric.

Marchand's proposal is articulated, rich and based on consistent criteria but it is applicable only to endocentric compounds. Marchand, in fact, maintains that what is known in the literature as exocentric compounds are to be analyzed as containing some sort of (categorizing) zero suffix and, as such, they are formations not pertaining to the compounding domain but to derivation.

Finally, the classification proposed by Bally seems to be based on a unique criterion, that of the grammatical relations holding between head and non-head constituents, viz. relations 'de coordination', 'd'accord' and 'de rection'. This proposal is consistent because it is based on a single criterion but, unfortunately, it is insufficient because of the absence of the notion «head». And, in some sense, we could say that Bally's and Fabb's proposals are complementary. A further problem with Bally's proposal, however, is the fact that included in the domain of compounds there are forms that today would be clearly considered as phrases (*maison de campagne*).

2. A new proposal

We propose a novel classification of compounds which is based on a very simple assumption. What is special about compounds is the fact that the two constituents are linked by a grammatical relation which is not overtly expressed (cf. *apron string* vs. *string of the apron*). Therefore, we would like to suggest that the classification of compounds be uniquely and consistently based on this criterion⁸. The possible grammatical relations holding between the two constituents of a compound are basically the relations that hold in syntactic constructions: subordination, coordination and attribution. The classification of compounds we propose is thus the following:



Compounds are classified as ‘subordinate’ whenever there is a ‘complement’ relation between the two constituents. In a compound such as *taxi driver*, *taxi* is clearly the «complement» of the deverbal head. We argue that this is also the case in compounds such as *apron string*, where *apron* is in an «of relation» with *string*. Complement relations, however, are not exclusively ‘of relations’: *apron string*, in fact, can have different interpretations: ‘string of an apron’, ‘string on a apron’, ‘string in a apron’, etc. Nevertheless, the relation between the two constituents is always a complement relation, namely a subordinative one.

⁸ This position is not completely new. For example Marchand (1969:18) observed that all compounds can be explained on the basis of the syntactic relations underlying the corresponding sentences. This observation is interesting, even though not workable within the lexicalist framework which rejects derivation of compounds from sentences.

Note that there is a clear subordination relation also when the head is not present, as in *cut throat* or in *lavapiatti* and not only when the compound is endocentric as *taxi driver*.

Attributive compounds are formed either by a noun and an adjective, as in *blue cheese* (where the adjective expresses a property and is in a modifier relation to the noun) or by two nouns, where the non-head very often is used somehow metaphorically, expressing an attribute of the head (cf. *snail mail*, *sword fish*).

Coordinate compounds are those formations whose constituents are tied by the conjunction «and». They are potentially recursive even in Romance languages (cf. It. *poeta pittore regista* 'poet-painter-director') where recursion in compounds is not usual. From a semantic point of view, such compounds can be considered as having two heads (*poet painter* is both a «poet» and a «painter»).

Compounds of these three classes can be both endocentric and exocentric. Therefore, all the compounds exemplified in (1) can be accommodated in six classes, as illustrated below:

3) SUBORDINATE		ATTRIB		COORDINATE		
endo	exo	endo	exo	endo	exo	
love story	loudmouth			bitter sweet		Bl.
steamboat		blackboard		oaktree	Austria-H.	March.
				girlfriend		
	pickpocket			fighter-bomber		
coffee cup			greybeard	learned-driver	mother-child	Sp.
lipstick	lavapiatti		greeneyed	poet doctor		Olsen
	lavapiatti	blackbird	Kiekkopf	poeta pintor	elun-ai	Hasp.
			Kahlkopf	woman doctor	Sch.H	Bauer
			auricomus	Fürstbischof	candra-ditya-u	Booij

Since the compounds in (3) are not representative of all possible types of compounds (note, for example that under the column 'attributive/endocentric' there are no compounds with a Noun as non-head) we will enrich the list with some other examples, as in (4):

(4)

SUBORDINATE	ATTRIB	COORDINATE
-------------	--------	------------

endo	exo ⁹	endo	exo	endo	exo
apple cake	kill joy	blue cheese	white collar	actor author	mind brain
brain death	cut throat	atomic bomb	green house	priest hermit	mother child
finger print		back yard	pale face	singer bassist	north east
mail man		French kiss	long legs	dancer singer	
sun glasses		ape man	free lance	artist designer	
water pipe		ghost writer		fighter bomber	
taxi driver		key word		king emperor	
stone cutter		public opinion		merchant tailor	
arm control		sword fish			
baby care					
agora phobia					

In a framework like the one we are proposing, neoclassical compounds - involving that particular kind of elements called «semi-words» (sW) (Scalise 1984) - are to be considered as subordinate. In this way, the use of a label which is only «descriptive» of the origin of constituents can be avoided. Subordination is, in fact, the relationship tying together the two constituents of these formations, as the examples in (5) show:

5) neoclassical compounds

sW+sW	sW+N	N+sW
calligraphy	apiculture	aerodrome
hydrology	hydrophobia	insecticide
philosophy	antropomorphism	parkingmeter

In fact, *hydrology* is the ‘science of water’, *hydrophobia* is the ‘fear of water’, and so on. Also the so-called phrasal compounds can be inserted into one of the proposed classes. Consider the following examples:

6) phrasal compounds

- [floor of a birdcage] taste
- [punch in the stomach] effect
- [pipe and slipper] husband

⁹ The low figures of exocentric compounds is intended to reflect the fact that in Germanic languages exocentricity is not so widespread as in Romance languages. This is probably due to the fact that Germanic languages are head final. In a Romance language such as Italian, on the contrary, the most productive type of compounds is the exocentric V+N compound (like *scacciapensieri*).

According to the criterion underlying our classification, these compounds have to be seen as belonging to the attributive class: in fact, the non-head phrases involved in these constructions function as properties qualifying the head nouns (*a terrible taste, a painful effect, a boring husband*). In these compounds the non head has a metaphorical interpretation: to say that something created a «punch in the stomach effect» on you means that you can feel like having received a punch in the stomach but you do not really need to actually have received it; and the meaning does not change if instead of «pipe and slippers» a husband is of the kind «newspapers and TV».

2.1. Selection

The distinction we draw among subordinate, coordinate and attributive compounds can be further supported by the modality by which the head selects the non head in each of these three classes (cf. Scalise, Bisetto, Guevara 2005). If we represent compounds making use (in a slightly modified way) of Lieber's (2004) theory of lexical semantics representation, according to which every lexeme is represented with a skeleton (containing grammatical information) and a body (containing encyclopaedic information), the following picture can be drawn (arrows indicate matching features; heads are underlined):

(7) selection in compounds

a) coordinate compounds

<u>actor</u>		<u>director</u>
[Thing[+com, -abst, +an]([X],[]),[Event ACT [X]]]	← --- ►	[Thing[+com, -abst, +an]([X],[]), [Event DIRECT [X]]]
<human, professional>	← --- ►	<human, professional>
<show business>	← --- ►	<show business>
<works in theatres, films, etc.>	← --- ►	<works in theatres, films, etc.>
<receives directions>		<gives directions>
< ... >		< ... >

b) N+N subordinate compounds

<u>apple</u>		<u>cake</u>
[Thing [+com, -abst, -an] ([])]		[Thing [+com, -abst, -an] ([])]
<physical>		<physical>

<shape>		<shape>
<edible>	←----→	<edible>
<can be an ingredient>	←----→	<made with ingredients>
<....>		<baked>
		<made for parties>

c) Attributive compounds

<i>snail</i>		<i>mail</i>
[Thing [+com, -abst, +an] ([])]		[Thing [+com, -abst, -an] ([])]
<gastropod>		<institution>
<secretes slime>		<means of communication>
<very slow>	←----→	<takes time>
<....>		<....>

Coordinate compounds (in 7a) are characterized by a virtual identity on both levels of representation: matching of the skeletons of the two constituents and a high level of matching features in the encyclopaedic body.

On the contrary, in subordinate compounds like in 7b), both endocentric and exocentric, the skeleton seems not to play a significant role: what really matters in this type of compound is the set of encyclopaedic features of the two bodies. At least one of the features of the head constituent must be matched by the encyclopaedic features characterizing the non-head constituent (such as <edible> and <made with ingredients> in the example above).

In attributive compounds with a noun as non-head the skeleton plays no significant role (as in subordinate compounds). What matters is that the non-head matches at least one of the encyclopaedic features of the head. The only information pertaining to the non-head that is present in the output is the matched feature: the rest of the information is ignored (for instance, the feature <gastropod> is not present at all in *snail mail*). The non-head has the sole function of specifying an attribute of the head's body (in this case «slowness»). In other words, the non-head is almost «adjectival»: in LCS terms, it is interpreted as a Property and not any longer as a Thing (two lexical-conceptual categories, related to, but not to be confused with, the syntactic categories Noun and Adjective, cf. Jackendoff 1990: 43–58). The attributive relation is obviously self-evident if the non-head is an Adjective (e.g. *blackboard*).

Different types of compounds have thus a different mechanism according to which the non head constituent is selected.

3. Some residual problems and some provisional conclusions

The proposed classification is generally clear and it seems to work properly. There are of course borderline cases where the analysis is more complicated. For example, compounds such as *greybeard* or *greeneyed* have been classified here as attributive (exocentric). Now, the attributive relation is indeed present between the two free constituents of this structure (*grey-beard*; *green eye*). But if one takes into consideration their whole structure, it can be observed that in *grey beard*, for example, the relationship between [grey beard] and the (non realized) external head is not the same. This is also true for *green eyed*, where besides the attributive relation between *green* and *eye*, there is another grammatical relation between the (realized) head *-ed* and *green eye*, probably a subordinative relationship. There is, then, a case where two different types of relationships are present in the same compound: which of the two do we consider primary?

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.

Attributive compounds, for example, do not constitute a homogeneous class because the non-head can be either an adjective (*blackbird*), or a noun (*sword fish*) or a verb (*play ground*). This means that the classification could be possibly enriched by a further subdivision based on categories.

It is also likely that within each of the three identified macrocategories there will be need for further semantic analysis in order to capture Marchand's intuition according to which what he calls copulative compounds can be further distinguished into, for example, subsumptive and additive compounds. In our classification, we do not distinguish among coordinate compounds (they all are linked by an «and» operator) but the semantic relation in subsumptive and additive compounds is clearly not identical (hyponymy in one case, synonymy in the other).

In conclusion, in this paper we have suggested that an adequate classification of compounds has to be done primarily and consistently on grammatical grounds. And exactly the grammatical relations between the constituents of a compound can allow an homogeneous grouping (at least on a first level) of compounds of different languages. However, we do not exclude (to the contrary) that there could be a second level of analysis, based on more subtle distinctions.

References

- Allen M.R. (1978), *Morphological Investigations*, PhD Dissertation, University of Connecticut.
- Bally, C. (1944), *Linguistique générale et linguistique française*, 2. ed. Berne, A. Francke.
- Bauer, L. (2001), *Compounding*, in Haspelmath, M. et al. (eds.), *Language Typology and Language Universals*, Vol. I, Berlin, Walter de Gruyter, 695-707.
- Bloomfield, L. (1933), *Language*, New York, Holt.
- Booij G. (2005), *The Grammar of Words*, Oxford, Oxford University Press.
- Fabb, N. (1998), *Compounding*, in Spencer A. and A. M. Zwicky (eds.) *Handbook of Morphology*, Oxford, Blackwell, 66-83.
- Hacken ten, P. (1994), *Defining Morphology. A Principled Approach to Determining the Boundaries of Compounding, Derivation, and Inflection*, Hildesheim, Olms.
- Haspelmath, M. (2002), *Understanding Morphology*, London, Arnold.
- Jackendoff, R. (1990), *Semantic Structures*, Cambridge (MA), MIT Press.
- Lieber, R. (1992), *Deconstructing Morphology*, Chicago-London, The University of Chicago Press.
- Lieber, R. (2004), *Morphology and Lexical Semantics*, Cambridge, Cambridge University Press.
- Marchand, H. (1969), *The Categories and Types of Present Day English Word Formation*, München, Beck'sche Verlagsbuchhandlung.
- Olsen, S. (2001), *Copulative Compounds. A Closer Look at the Interface Between Morphology and Syntax*, in «Yearbook of Morphology 2000», 279-320.
- Scalise S. (1984), *Generative Morphology*, Dordrecht, Foris.
- Scalise S. (ed.) (1992), *The Morphology of Compounding*, special issue of «Rivista di Linguistica», 4 (1).

Scalise S., A. Bisetto and E. Guevara (2005), *Selection in Compounding and Derivation*, in Dressler W. U., D. et al. (eds.), *Morphology and its Demarcation*, Amsterdam/Philadelphia, John Benjamins Publishing Company, 133-50.

Selkirk, E. (1982), *The Syntax of Words*, Cambridge (MA), MIT Press.

Spencer, A. (1991), *Morphological Theory*, Oxford, Blackwell.

SUMMARY

In questo articolo si discute della classificazione dei composti a partire dalle diverse analisi che ne sono state proposte, in particolare in questi ultimi anni. I problemi che le classificazioni correnti pongono sono sia di terminologia che di coerenza di criteri classificatori. In questo lavoro si propone una classificazione «di primo livello» che utilizzi un criterio unico, quello del rapporto grammaticale tra costituenti. In questo modo vengono identificati tre macroclassi di composti: subordinati, coordinati ed attributivi. Le nozioni di endocentricità ed esocentricità, abitualmente considerate come criteri di primo livello, sono in realtà subordinate rispetto al criterio grammaticale.