3 Compounding

NIGEL FABB

1 Overview

A compound is a word which consists of two or more words. For example, the Malay compound *mata-hari* 'sun' is a word which consists of two words: *mata* 'eye' and *hari* 'day'. Compounds are subject to phonological and morphological processes, which may be specific to compounds or may be shared with other structures, whether derived words or phrases; we explore some of these, and their implications, in this chapter. The words in a compound retain a meaning similar to their meaning as isolated words, but with certain restrictions; for example, a noun in a compound will have a generic rather than a referential function: as Downing (1977) puts it, not every man who takes out the garbage is a *garbage man*.

1.1 Structure and interpretation

The meaning of a compound is usually to some extent compositional, though it is often not predictable. For example, *popcorn* is a kind of corn which pops; once you know the meaning, it is possible to see how the parts contribute to the whole – but if you do not know the meaning of the whole, you are not certain to guess it by looking at the meaning of the parts. This lack of predictability arises mainly from two characteristics of compounds: (a) compounds are subject to processes of semantic drift, which can include metonymy, so that a *redhead* is a person who has red hair; (b) there are many possible semantic relations between the parts in a compound, as between the parts in a sentence, but unlike a sentence, in a compound, case, prepositions and structural position are not available to clarify the semantic relation.

1.1.1 Endocentric and exocentric compounds Compounds which have a head are called 'endocentric compounds'. A head of a compound has similar

characteristics to the head of a phrase: it represents the core meaning of the constituent, and it is of the same word class. For example, in *sneak-thief*, *thief* is the head (a sneak-thief is a kind of thief; *thief* and *sneak-thief* are both nouns). Compounds without a head are called 'exocentric compounds' or 'bahuvrihi compounds' (the Sanskrit name). The distinction between endocentric and exocentric compounds is sometimes a matter of interpretation, and is often of little relevance; for example, whether you think *greenhouse* is an endocentric or exocentric compound depends on whether you think it is a kind of house. The major interest in the head of a compound relates to the fact that where there is a clear head, its position seems to be constrained; endocentric compounds tend to have heads in a language systematically on either the right (e.g. English) or left (e.g. Vietnamese, French).

1.1.2 Co-ordinate compounds There is a third kind of compound, where there is some reason to think of both words as equally sharing head-like characteristics, as in *student-prince* (both a student and a prince); these are called 'appositional' or 'co-ordinate' or 'dvandva' (the Sanskrit name) compounds. Co-ordinate compounds can be a combination of synonyms (example from Haitian):

toro-bèf (*bull-cow*) 'male cow'

a combination of antonyms (example from French):

aigre-doux (*sour-sweet*)

or a combination of parallel things (example from Malayalam):

acchanammamaaṟə (*father-mother-pl.*) 'parents'

Co-ordinate compounds can have special characteristics in a language. For example, in Mandarin, co-ordinate compounds behave differently in terms of theta-assignment from (endocentric) resultative verb compounds. And in Malayalam, co-ordinate compounds are not affected by gemination processes which other compounds undergo (see section 5.2).

1.1.3 The semantic relations between the parts The semantic relations between the parts of a compound can often be understood in terms of modification; this is true even for some exocentric compounds like *redhead*. Modifier–modifiee relations are often found in compounds which resemble equivalent phrases; this is true, for example, of English AN%N[1] compounds and many Mandarin compounds (see Anderson 1985b). It is not always the case, though; the French compound *est-allemand* (East German) corresponds to a phrase *allemand de l'est* (German from the East). In addition, many compounds manifest relations which can be interpreted as predicator–argument relations, as in

sunrise or *pull-chain*. Note that in *pull-chain*, *chain* can be interpreted as an argument of *pull*, and at the same time *pull* can be interpreted as a modifier of *chain*.

1.1.4 Transparency: interpretive and formal The transparency and predictability of a compound are sometimes correlated with its structural transparency. For example, in languages with two distinct types of compound where one is more interpretively transparent than the other, the less interpretively transparent type will often be subject to greater phonological or morphological modification. A diachronic loss of transparency (both formal and interpretive) can be seen in the process whereby a part of a compound becomes an affix, as in the development of English *-like* as the second part of a compound to become the derivational suffix *-ly*.

1.2 Types of compound

Accounts of compounds have divided them into classes. Some of these – such as the exocentric, endocentric and appositional types, or the various interpretive types (modifier–modifiee, complement–predicator, etc.) – are widespread across languages. Then there are compound types which are language- or language-family-specific, such as the Japanese postsyntactic compounds (4.3), Hebrew construct state nominals (4.2), or Mandarin resultative verb compounds (4.1). Other types of compounds are found intermittently; these include synthetic compounds (section 1.2.1), incorporation compounds (section 1.2.2) and reduplication compounds (section 1.2.3).

1.2.1 Synthetic (verbal) compounds The synthetic compound (also called 'verbal compound') is characterized by a co-occurrence of particular formal characteristics with particular restrictions on interpretation. Not all languages have synthetic compounds (e.g. English does, but French does not). The formal characteristic is that a synthetic compound has as its head a derived word consisting of a verb plus one of a set of affixes (many writers on English restrict this to agentive *-er*, nominal and adjectival *-ing*, and the passive adjectival *-en*). Thus the following are formally characterized as synthetic compounds:

expert-test-ed
checker-play-ing (as an adjective: a checker-playing king)
window-clean-ing (as a noun)
meat-eat-er

(There is some disagreement about whether other affixes should also be included, so that *slum-clear-ance* for example would be a synthetic compound.) Compounds with this structure are subject to various restrictions (summarized by Roeper and Siegel 1978), most prominent of which is that the left-hand member must be interpreted as equivalent to a syntactic 'first sister' of the

right-hand member. We discuss this in section 3.2. There have been many accounts of synthetic compounds, including Roeper and Siegel 1978, Selkirk 1982, Lieber 1983, Fabb 1984, Botha 1981 on Afrikaans, and Brousseau 1988 on Fon; see also section 2.1.2. Many of the relevant arguments are summarized in Spencer 1991.

1.2.2 Incorporation compounds In some languages, incorporation words resemble compounds: for example, both a verb and an incorporated noun may exist as independent words. Bybee (1985) comments on this, but suggests that even where the two parts may be independently attested words, an incorporation word may differ from a compound in certain ways. This includes phonological or morphological differences between incorporated and free forms of a word. (In fact, as we have seen, this is true also of some compound processes; e.g. segment loss, which is found in Tiwi incorporation words, is found also in ASL compounds.) Another difference which Bybee suggests may distinguish incorporation from compounding processes is that the incorporation of a word may depend on its semantic class. For example, in Pawnee it is mainly body part words which are incorporated, while various kinds of name are not (such as personal names, kinship terms, names of particular species of tree, etc.). It is possible that compounding is not restricted in this way; as we will see, semantic restrictions on compounding tend to be in terms of the relation between the parts rather than in terms of the individual meanings of the parts.

1.2.3 Repetition compounds Whole-word reduplication is sometimes described as a compounding process, because each part of the resulting word corresponds to an independently attested word. Steever (1988), for example, describes as compounds Tamil words which are generated through reduplication: *vantu* 'coming' is reduplicated as *vantu-vantu* 'coming time and again'. In another type of Tamil reduplicated compound, the second word is slightly modified: *viyāparam* 'business' becomes *viyāparam-kiyāparam* 'business and such'. English examples of this type of compound include words like *higgledy-piggledy*, *hotchpotch*, and so on.

1.3 Compounds which contain 'bound words'

In a prototypical compound, both parts are independently attested as words. However, it is possible to find words which can be parsed into an independently attested word plus another morpheme which is not an independently attested word but also does not appear to be an affix (see Aronoff 1976). Here are some examples from English (unattested part italicized):

church-*goer* iron*monger* *tele*vision *cran*berry

The part which is not attested as an isolated word is sometimes found in other words as well; in some cases it may *become* an attested word: for example, *telly*

(= television). These parts fail to resemble affixes morphologically (they are relatively unproductive compared to most affixes), and there is no good evidence on phonological grounds for considering them to be affixes. They are also unlike affixes semantically; judging by their contribution to the word's meaning, they have lexical rather than grammatical meanings. A similar example from standard Tamil (Steever 1988) is the form *nāḷ* 'goodness', which is not independently attested, but can be found in compound nouns such as *nalla-nāḷ* 'good-day'.

2 The structure of compounds

One thing that is reasonably clear about the structure of compounds is that they contain two words, and the distinctness of these two components is visible to various rules. Other aspects of structure are not so obvious, however: in the order of the parts regulated by rules 2.1, are their word classes visible to any rules 2.2, and are three-word compounds hierarchically structured (2.3)?

2.1 Directionality

A compound can be 'directional' in two senses. One sense involves the position of the head: whether on the right or the left. The other sense involves the direction of the relation between the parts of the compound: the direction of modification in a noun–noun compound (e.g. in *log cabin* modification is rightwards) or the direction of complementation in a verb-based compound (e.g. in *push-bike* complementation is rightwards). Notice that the two senses of directionality can be independent, because a compound can have internal modification or complementation without having a head: *killjoy* has no head, but it does have a predicator–complement order. This is an important descriptive issue; some accounts assume that a modifier–modifee or predicator–argument relation inside a compound is itself evidence that part of the compound is a head. To the extent that there are any useful claims about directionality of the head to be made, it is probably best to focus on the narrowest definition of head (which involves a semantic link between head and whole); see Brousseau 1988 for arguments to this effect.

2.1.1 The location of the head In English, the head of an endocentric word is on the right. In French, the head is on the left (as in *bal masqué*, 'masked ball'). It has been argued (e.g. by DiSciullo and Williams 1987) that all true endocentric compounds are right-headed, and that any left-headed compounds should be considered as exceptions – for example, as phrases which have been reanalysed as words. This is not a widely accepted account (it is based on the controversial Right-Hand Head rule of Williams 1981b).

2.1.2 An argument about directionality and synthetic compounds

Brousseau suggests that the direction of relations inside a compound is responsible for determining whether a language is able to have synthetic compounds. She suggests that in a synthetic compound, the non-head must modify the head, and be a complement of the head. This means that the direction of modification must be opposite to the direction of complementation:

—modifies→
meat-eater
←takes as complement—

Because in English, the directions of complement taking and modification are opposed, synthetic compounds are possible (the same is true of Fon). In French and Haitian (a Fon–French creole), on the other hand, both modification and complementation are leftward, and so the particular conditions which allow a synthetic compound are absent – hence there are no synthetic compounds in these languages.

2.2 The word classes of the component words

Another question which must be asked about the structure of compounds is whether the class of the component words is relevant, or whether word class is lost when the words are formed into a compound. 'Relevant' would mean, for example, visibility of word class to a class-sensitive phonological or morphological rule. Little attention has been focused on this question; clearly some compound-internal affixation rules are class-sensitive (but this might be because they are added before the compound is formed).

Most of the attention to word class in a compound has focused on the attested word-class structures of compounds in a language. For example, in Punjabi (Akhtar 1992) there are large numbers of compounds involving a combination NN, AN, AA, NV and VV, but none with a structure VA, and very few with a structure VN. Selkirk (1982) suggests that these facts about a language are best expressed by compound-specific rewriting rules analogous to phrase-structure rules. This approach has been adopted by many people, and is useful as a descriptive device. However, there are some fundamental differences between structure-building rules for compounds and structure-building rules for phrases:

(a) There is no true equivalent of X-bar theory as a constraint on compound-building rules. Most obviously, compounds need not have a head. More generally, it is hard to find structural generalizations across compound structures analogous to the generalizations expressed by X-bar theory for phrases.

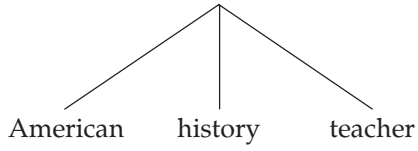
(b) Compound-building rules would rarely be recursive. In English, for example, the only clearly recursive type is the NN%N combination.

(c) There is a problem about productivity. Phrase-structure rules are fully productive; each rule can underlie an infinite number of phrases (partly because of recursion). But some rules for building compounds are manifested by very few actual compounds. Selkirk recognizes this, and distinguishes rule-built compounds from non-rule-built compounds. Perhaps, though, there is an alternative way of explaining the prevalence of certain compound types along functional rather than formal lines. Thus, in English, the prevalence of NN%N and AN%N types might be because of a functional need for compound nouns before other word classes, and because these have a modifier–modifee structure which is easily interpreted. This is a complex problem which requires metatheoretical decisions about the place of functional considerations and the meaning of productivity.

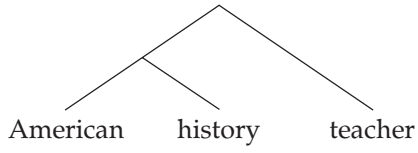
2.3 *Subconstituency in three (or more) word compounds*

Where compounds consist of three or more words, the compound can sometimes be interpreted by breaking it down into subconstituents. This is true, for example, of *chicken-leg-dinner*, which is interpreted by taking *chicken-leg* as a subcompound within the larger compound. In some cases, ambiguity arises from the possibility of two alternative groupings of words, as in *American history teacher* (a history teacher who is American or a teacher of American history). This fact about interpretation raises the question of whether three-or-more-word compounds might perhaps have a subconstituent (hierarchical) structure like a phrase (i.e. (b) or (c) rather than the flat structure (a)).

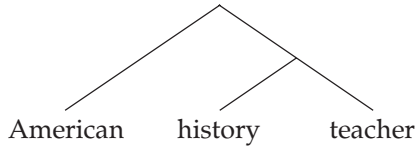
(a)



(b)



(c)



It is not obvious that the interpretive facts alone demonstrate the presence of a complex structure. The interpretive rules might simply pick any pair of adjacent units and make them into a unit, taking a 'syntactic' structure like (a) and building a 'semantic' structure like (b) or (c). Compare hierarchical *phrase* structure, which pre-exists any interpretive strategy, as can be shown by the sensitivity of syntactic processes (such as binding theory) to constituent structure. But compounds are relatively inert compared to syntactic constituents (no movement, anaphoric coindexing, etc.), so it is harder to find supporting evidence for complex constituent structure.

As we will see (section 5.1.1), there is some evidence in English from the stressing of four-word compounds that the compound-specific stress rules are sensitive to a subconstituent structure, and hence that such structure exists outside the interpretive component. Turkish also presents evidence for subconstituent structure in hierarchically interpreted compounds (Spencer 1991). The 'indefinite izafet compound' has the structure Noun + Noun-poss. when it consists of two words. The three-part compound has one of two types of structure, each corresponding to a different interpretation of its hierarchical structure (as bracketed):

[Noun + Noun-poss.] + Noun-poss.
Noun + [Noun + Noun-poss.]

Here the placement of the possessive affix is sensitive to hierarchical structure.

Most of the multiple-word compounds which are cited (particularly those cited as hierarchical compounds) tend to be combinations of nouns. This suggests that if there are structure-building rules for compounds, as discussed in the previous section, they are not generally recursive.

Note that sometimes a compound can be interpreted as having an internal subconstituency, without that subconstituent being an attested word. An example from Fon (Brousseau 1988) is:

gàn-gɔ̀-kɛ̀n NNN%N money-stomach-cord = money belt

This compound is presumably interpreted as N[NN], but the [NN] subconstituent does not exist as an independent word. This is not necessarily a strong argument against a hierarchical structure in this case, however, because there are also many examples of two-word compounds where one 'word' is not independently attested.

Note that some multiple-word compounds are not interpreted as having a hierarchical structure; this is true particularly of dvandva compounds such as the following from Tamil:

vīra-tīra-cakacaṅ-kaḷ (*courage-bravery-valour-pl.*) 'courage, bravery and valour'

An example of a compound which mixes non-hierarchical with hierarchical interpretation is the following, from Mandarin:

jī-yā-yú-ròu = 'animal foodstuffs' (i.e. chicken-duck-fish-meat)

Interpretively the structure is [[chicken + duck + fish] meat].

2.3.1 Hierarchies which include non-word components A three-member compound need not contain three words. For example, a morpheme may appear between two words in a compound or at one end of a compound. It has been argued by a number of writers that in (English) synthetic compounds, the suffix is a third constituent of the compound; synthetic compounds on this analysis have a hierarchical structure such as [[meat-eat]-er].

3 The interpretation of compounds: interpretive gaps

Extensive descriptive work has been undertaken on the semantic relations holding between the components of English compounds (see e.g. Levi 1978, Warren 1978). An interesting theme that arises from some of this work is the possibility that there are certain gaps: semantic relations between the parts of a word which are possible in principle, but are not attested in practice. In this section we look at two such gaps in English compounds.

3.1 The missing goal

One of the commonest kinds of compound in English is the NN%N type. Many different relations can be interpreted as holding between the two members of such a compound. The interesting question in these cases is whether any relationships are not attested; in surveys of NN%N compounds, both Downing (1977) and Warren (1978) found that while source (something moved away from) was attested, goal (something moved towards) was only marginally attested. Warren found only fourteen potential examples out of 3,994 compounds, and suggests that these may not even be true examples of 'goal' compounds. In her list of relations between the parts of nominal compounds, Levi (1978) has 'from' (e.g. *store-clothes*) but not 'to'.

The same gap can be seen in other compounds; for example, while we find VN%N compounds like *print-shop* (a shop where printing takes place), there are no compounds like *go-place* (meaning a place to which someone goes). This gap also appears in synthetic compounds: *heaven-sent* can be interpreted only as sent *from* heaven, not sent *to* heaven. Note that apparent goal compounds like *church-goer* actually mean 'someone who attends church' (not someone who moves towards church); similarly, *sea-going* means 'going on the sea' (not *to* the sea).

3.2 Synthetic compounds in English: the absence of 'subject'

In a synthetic compound, the crucial interpretive restriction is that the left-hand word (a noun, adverb or adjective) must be interpretable as a complement of the right-hand word (and must *not* be interpretable as an external argument or subject). In effect, synthetic compounds with *-ing* or *-er* are like reversed active verb phrases with equivalent components (play checkers > checker-playing), while synthetic compounds with passive *-en* are like reversed passive verb phrases (tested by experts > expert tested). Synthetic compounds thus differ from other compounds (sometimes called 'root compounds'); hence while **bird-singing* is excluded, there is a compound *bird-song* where the left-hand member is interpretable as the subject of the right-hand member.

Synthetic compounds are interesting because the rules for their interpretation seem to be related to rules for building the meaning of sentences (e.g. the assignment of thematic roles to particular positions in a sentence, depending on the active or passive nature of the verb).

3.3 Ways of explaining interpretive gaps

Interpretive gaps in compounds could in principle be explained in one of three ways:

(a) Constrain the compound-building rules to make them sensitive to interpretation-relevant aspects such as thematic relations. Levi (1978) takes this approach within a Generative Semantics framework, for compound nouns. This is also the approach taken by Roeper and Siegel (1978) in an Extended Standard Theory framework; they build synthetic compounds by a special (transformational) rule which takes a combination of a verb and a subcategorized complement as input. The gaps exist because there is no possibility of building synthetic compounds which have a verb combined with a non-subcategorized argument such as its subject. But Roeper and Siegel's approach (in particular) runs into a problem. Consider, for example, the compound *bird-singing*. This cannot be built by the synthetic-compound-building rule, because it combines the verb with a non-subcategorized argument (its subject). But nothing stops it being built by an alternative rule – the root-compound-building rule, which takes two nouns and combines them (*bird + singing*). This rule is not subject to thematic constraints, as can be seen in subject–predicate compounds like *sunrise*. So whatever rules out *bird-singing* as a root compound is clearly not associated with the compound-building rules.

(b) Instead, it may be that the subject is ruled out by some filter which looks at the compound, and if it has the structure of a synthetic compound, applies certain constraints on interpretation to it. This would differentiate root from synthetic compounds not in how they are built, but in their surface form.

(c) A third possible approach would be to explain interpretive gaps in compounds in terms of general constraints on the possible meanings of a word (see Carter 1976). Such an approach does not necessarily require that compound-internal thematic relations be specified, because these relations are not explicitly referred to. This approach might provide an explanation of the 'goal' gap. Note that this gap is found also in non-compound word formation: Hale and Keyser (1992) point out that while there is a verb *shelve*, meaning 'put on a shelf', there is no verb *church*, meaning 'go to church'. It may be that a meaning of 'movement towards' is incompatible with some aspect of possible word meaning. For example, Downing (1977) comments that 'unambiguously fortuitous or temporary relationships' are ruled out in favour of generic or habitual relationships. Perhaps 'movement towards' is ruled out in general because it is not usually a generic or habitual relationship: in this light, it is interesting to compare 'source' (movement from) with 'goal' (movement towards). The source of something remains a stable and permanent property of that thing; the goal of something is its goal only while it is travelling towards it.

4 Compounds and syntax

In this section we look at some language (or language-family)-specific compound types which have an internal structure open to syntactic manipulation and visible to syntactic processes. Incorporation compounds are a clear example, and synthetic compounds have also been argued to have such a structure (e.g. by Fabb 1988). There are two complicating factors when considering the 'syntactic' aspects of compounds.

One is that compounds tend to have relatively fixed meanings, so that it is difficult, for example, to modify them; the question of syntax vs morphology may be irrelevant here. Thus, for example, the ASL compound 'blue-spot' (= bruise) cannot be morphologically modified to '*darkblue-spot' (?= bad bruise); and the French compound *garde-malade* (= nurse) cannot be syntactically modified to *garde-bien-malade* (?= good nurse).

The second problem relates to the possibility that some compounds are the result of lexicalization of phrases. Thus, while in English it is generally impossible to have *the* inside a compound, there is a word *middle of the road* which looks like a compound but may best be analysed as a lexicalized phrase; the same can be said for many French compounds such as *pomme-de-terre* (= potato) or *trompe-l'œil* (= illusion), both of which contain typically syntactic components, the preposition *de* or the article *l'*. Compounds containing *and*, such as *foot-and-mouth disease* may perhaps be dealt with by claiming lexicalization of a phrase; or it may simply be that co-ordination can involve parts of words with no syntactic implications (see Sproat 1985 and Bates 1988 for different positions on this).

4.1 Mandarin resultative verb compounds

Y-F. Li (1990) discusses resultative verb compounds in Mandarin, which have a VV%V structure:

Baoyu qu-lei-le ma
Baoyu ride-tired-asp horse
'Baoyu rode a horse (and as a result it/Baoyu got) tired'

The resultative verb compound here is *qu-lei*. The first verb describes an action 'ride'; the second verb characterizes the result of that action as a state, 'tired', of the compound verb's subject (Baoyu) or object (that horse). The compound verb thus inherits theta-role-assigning properties from the component verbs. This is one reason for thinking that the internal structure of the compound may be visible in the syntax. Another reason comes from the fact that the first verb can be duplicated and followed by the object, so that the above sentence has this as an alternative form:

Baoyu qu ma qu-lei-le

A third reason for assuming syntactic visibility is that it is possible to modify the compound by inserting *de* or *bu* between the two verbs; the particles add the meaning 'can' or 'cannot' respectively (see Anderson 1985b).

Li suggests that the identification of the first verb as the head of the resultative verb compound has important consequences. He compares the resultative verb compound with a co-ordinate VV%V compound type. In the resultative verb compound, only the first verb must assign its external argument to the subject of the sentence, because only the first verb is the head of the compound word; the second word is free to assign its external argument to the subject or the object in the sentence, resulting in the ambiguity illustrated above. In a co-ordinate compound, however, both verbs must assign their external argument to the subject of the sentence, because they are both heads of the compound word (hence the ambiguities found in the resultative verb compound are not available to co-ordinate compounds).

4.2 Hebrew construct state nominals

Borer (1988) distinguishes two kinds of multi-word words in Hebrew, construct state nominals:

ca9if ha-yaldá (*scarf the-girl*) 'the girl's scarf'

and compounds:

beyt xolím (*house sicks*) 'hospital'

The two kinds of word can be seen as two kinds of compound. Both kinds of compound have word-like phonology (e.g. both have a single primary stress) and morphology (e.g. both are pluralized in the same way). Neither compound can be extracted from. The difference between them is that construct state nominals are more transparent interpretively and syntactically than other compounds; for example, definiteness and plurality on the components of a construct state nominal are taken into account by sentence-interpretation rules. Borer argues that the distinction should be achieved by having word-formation rules in the lexicon (to build ordinary compounds) or in the syntax (to build construct state nominals). A word built in the syntax will have its components visible to syntactic processes.

4.3 Japanese post-syntactic compounds

Shibatani and Kageyama (1988) argue that Japanese has lexical and postsyntactic compounds. A postsyntactic compound is 'postsyntactic' because it appears to consist of the turning of syntactic material into a compound: it consists of a noun phrase followed by a verbal noun which takes it as a complement (i.e. the compound is subject to a kind of first-sister constraint, and so cannot take a scrambled sentence as input). Here is an example of such a compound:

zikken-syuuryoo-go
experiment-finish-after = 'after the experiment was completed'

Postsyntactic compounds are interpretively transparent. They are also transparent to syntactic processes such as anaphoric co-reference, and they can contain demonstratives and honorifics. They do not have the tonal pattern associated with lexical words. Despite these syntactic characteristics, Shibatani and Kageyama argue that the constituents as a whole are words (i.e. compounds), because the case particle (following the NP) is omitted, there is no tense on the verbal element, and the compound cannot be interrupted by an adverb. Like Borer, they thus allow compounds to be built at different levels (in this case, lexicon and PF).

5 Phonological and morphological processes and compounds

Compounds resemble derived words in some ways, and resemble phrasal combinations in other ways. This is true also of the phonological and morphological processes which apply to compounds. For example, some phonological processes group compounds with derived words; others group compounds

with phrases; and others isolate compounds as a distinct class. Most recent discussions of these groupings have been in a Lexical Morphology/Phonology framework, and have been used to argue for a level-ordering of morphological and phonological processes (see section 5.2 for an illustration).

5.1 Suprasegmental processes

5.1.1 Stress One of the most commonly cited compound-specific phonological rules is stress assignment. Compounds may be subject to a rule which places heavier stress on one word. In English, for example, the compound *big-foot* (the name of a monster) is differentiated from the phrase *big foot* by the heavier stressing of *big* relative to *foot* in the compound. (In English and Danish, the first word is stressed; in Italian and Spanish, the second word.) Even where there is not a compound-specific stress rule, compounds can still show interesting stress properties. For example, Anderson (1985b) shows that in Mandarin, contrastive stress 'sees' a compound as a single word, so there is only one place in the compound where contrastive stress can be placed; by comparison, a two-word phrase can have contrastive stress on either word.

The stress pattern of English compounds has been the source of extensive analysis. It is relevant primarily in three areas:

(a) The stress pattern of compounds may indicate the presence of hierarchical structure inside the compound. Since most versions of the compound stress rule apply to two-word compounds, it is interesting to see what happens to the stress in larger compounds. In these, there seems to be clear evidence that stress assignment recognizes internal subcomponents; for example, in a four-word compound like *student essay récord book*, the greatest stress is on *record* because a rule recognizes the presence of a subconstituent *record book* and places extra stress here. Note that a different hierarchical structure (corresponding to a different hierarchy of interpretation) involves a different stress pattern, as in *American hístory teacher association*. This is one of the few pieces of evidence that multi-word compounds in English have a complex internal formal structure.

(b) Compounds must be handled by some system of rules for stress assignment in English. Accounts vary in the extent to which they have rules specific to compounds. Chomsky and Halle (1968) have a separate Compound Rule. Breaking with this tradition, Liberman and Prince (1977) adapt the Lexical Category Prominence Rule so that it applies to derived words and compounds; similarly Halle and Vergnaud (1987) draw attention to the close similarity between rules for compounds and rules for derived words. The question of which rules apply to stress in compounds is relevant to questions of ordering in the lexicon (e.g. level-ordering). This extends beyond English: Shaw (1985) shows that the Dakota 'syntactic compounds' have two major stresses, while the 'lexical compounds' have one major stress; she explains this by having the

compounds formed at different levels in a level-ordered lexicon, with stress rules also level-ordered.

(c) English compounds do not all have initial stress; examples of level- and final-stressed compounds are *knee-déep* and *apprentice-wélder* respectively. This opens up the possibility that there are structural (or interpretive) characteristics of compounds, which stress rules are sensitive to. Bates (1988), for example, develops an analysis in which the English compound stress rules are sensitive to whether an initial adjective is predicative (initial stress) or not (final stress). Analyses along these lines provide potential evidence for internal structure in a compound (evidence which is otherwise sparse, see section 2.3). Note that the initial/final stress issue is in some cases a matter of dialectal difference; compare British English *ice-créam* and *hot dóg* with American English *íce-cream* and *hót dog*.

5.1.2 Other suprasegmental processes While stress is the most commonly cited compound-specific suprasegmental process, other such processes may differentiate compounds from derived words or phrasal combinations. In many Mandarin Chinese compounds the second element loses stress because it first loses its tone (Anderson 1985b). In Fon, nasalization occurs in a suffix, but not in the second word in a compound (Brousseau 1988). And Klima and Bellugi (1979) argue that in American Sign Language, temporal rhythm is analogous to stress; a compound word and a simple word take more or less the same amount of time to produce, because in a compound word, the first word is made significantly more quickly. Hence there is a compound-specific temporal reduction rule (cited in Liddell and Johnson 1986).

5.2 Phonological processes between the two words

Segmental phonological processes will tend to happen between two morphemes – either a word and an affix, or two words. There are examples of such processes which distinguish compounds from other combinations. In English, some phonological rules apply within derived words but not within compounds; M. Allen (1978) shows this for *beauty-parlour* vs *beauti-ful*. In Dakota (Shaw 1985) a different alliance can be found, with compounds patterning with derived words rather than phrases: 'lexical compounds' are subject to an epenthesis rule adding *a* to the end of the first word, which then feeds a continuant voicing rule; this rule is found also in derived words, but not in 'syntactic compounds'.

Malayalam (Mohanan 1982) provides a particularly complex picture of the visibility of compound structures to different phonological processes. Stem-final nasal deletion and stem-final vowel lengthening apply just to compounds. Other rules apply to both compounds and derived words (but not to sequences of words in a phrase), such as a vowel sandhi rule where two adjacent vowels are merged into one. A further complexity is that some rules apply only to a

subset of the compounds: stem-final and stem-initial gemination of obstruents in Dravidian stems occurs in compounds with a modifier–modifiee structure ('subcompounds') but not in co-ordinate compounds. Mohanan uses these facts as evidence for a theory of level-ordered lexical phonology/morphology, along the following lines:

stratum	morphological processes	vowel sandhi	vowel lengthening	nasal deletion	gemination
1	derivation	yes			
2	subcompounding	yes	yes	yes	yes
3	co-compounding	yes	yes	yes	
4	inflection				

5.3 Segment loss

American Sign Language compounds tend to undergo more radical formal restructuring than most spoken language compounds, with loss of segments being a common feature of compound formation. For example, the compound translated as 'black-name' (meaning bad reputation) involves loss of one of the segments of the word for 'name' (which on its own consists of a sequence of two identical segments). Liddell and Johnson (1986) comment that in ASL 'the average compound has the same number of segments as the average simple sign'.

5.4 Morphemes and compounds

5.4.1 Compound-specific morphemes In some languages, a morpheme (with no independent meaning) may be inserted between the two words. This morpheme may bear a historical relation to some affix, but is synchronically found only in compounds. In German, for example, *s* or *en* may be inserted:

Schwan-en-gesang 'swan song'

5.4.2 Inflectional morphology There have been two lines of research focusing on whether an inflectional morpheme (e.g. marking plurality, case or tense) can be found on a component word inside a compound.

One approach (e.g. Kiparsky 1982b) focusing on the question of level-ordering of morphological processes, looks at the distinction between regular and irregular inflectional morphology, and asks whether irregular morphology is more likely to appear on a compound-internal word (because it precedes the compounding process) than regular morphology (which comes later than the compounding process). The evidence from English is not particularly clear. As

predicted, regular inflection is lacking in some compounds like *footprints* (compare **foots-prints*), where irregular inflection is present in corresponding compounds like *teeth-marks*. But, *contra* predictions, sometimes regular inflection is present as in *arms race* where irregular inflection is missing as in *out-putted*.

The second line of enquiry focuses on the implications of having any inflection on a part of a compound, and whether this means that the internal structure of compounds is visible to the syntax (the answer to this depends on one's theory of inflection). For further discussion, see Anderson 1992: ch. 11. Questions about the visibility of inflection relate also to questions about the visibility of theta-assigning properties of subparts (as e.g. in Mandarin resultative compounds) and what this implies about the visibility of compound structure in syntax.

5.4.3 Derivational processes and compounds Compounds on the whole tend not to undergo derivational processes. However, there are some fairly clear examples such as English *bowler-hatted*, which can only be interpreted as *bowler-hat* + *ed*. It may be that English synthetic compounds are the result either of a verb-final compound being affixed with *-ing*, *-er*, etc., or of a combination of derivation and affixation (as in Roeper and Siegel's (1978) compound formation rule). There has also been discussion of compounds such as *transformational grammarian*, which on interpretive grounds would seem to involve suffixation of *transformational grammar*; there are, however, other ways of dealing with these 'bracketing paradoxes' (see Spencer 1988b, e.g.). Liddell and Johnson (1986) show that in ASL a reduplication process adds the meaning of 'regularly' to a verb; this process reduplicates all of a compound verb (though this may be an inflectional process).

5.4.4 Clitics and compounds Dakota postverbal clitics are attached in the lexicon (Shaw 1985). They attach to the first member of syntactic compounds, but not to the first member of lexical compounds.

6 Conclusion

Comparatively little theoretical work has been done on compounds. This is because compounds tend to be less phonologically or morphologically active than derived words, and less syntactically active than phrases: compounds are relatively inert. My view is that the three most interesting questions about compounds are as follows:

(a) If there are compound-building rules, what form do they take? Are they recursive (i.e. building hierarchical structure), and do they take word class into account? Are compound-building rules based on a universally available rule type (i.e. is 'compound' a universal?).

(b) How do compound-building rules interact with other rules – or, put slightly differently, what aspects of compound structure are visible to what other processes, whether syntactic, morphological or phonological? It is particularly important here to show that a certain structural fact about a compound is visible to two distinct rules (e.g. Bates (1988) argues that compound stress rules and interpretive rules both 'see' that a particular modifier in a compound is predicative rather than attributive). The question of the interaction of compound-building rules with other rules has been discussed particularly in the level-ordered systems of lexical phonology and morphology.

(c) Are there interpretive gaps? Are some sorts of meaning not creatable by putting compounds together – and if so, is this because of compound-specific rules, or because of more general principles of possible word meaning?

NOTE

1 XY%Z is to be interpreted as: [XY] is a compound of word class Z.