

The item **mousepad** is a compound: a **central, typical, prototypical** compound in several respects.

Prototypical compounds...

Topic		Refinements
1.1 ←	...are lexicalized – but not all are	→ 1.2
2.1 ←	<i>Most</i> compounds are nouns – but not all are	→ 2.2
3.1 ←	... consist of major lexical categories – but not all do	→ 3.2
4.1 ←	... consist of words, i.e. lexemes – but not all do	→ 4.2
5.1 ←	... have no internal inflections – but some do	→ 5.2
6.1 ←	... are binary (have two members) – but not all are	→ 6.2
7.1 ←	... are binary even if they have more layers – but not all are	→ 7.2
8.1 ←	... have no interfix – but some do	→ 8.2
9.1 ←	... may apply recursively – but not all do	→ 9.2
10.1 ←	... are endocentric (have a semantic head) – but not all	→ 10.2
11.1 ←	... are endocentric (have a structural head) – but not all	→ 11.2
12.1 ←	... are right-headed – but not all are (even in E)	→ 12.2
13.1 ←	... are subordinate –but not all compounds are subordinate	→ 13.2
14.1 ←	... tend to be left-branching (if containing more binary branches) but – not all are	→ 14.2
15.1 ←	... tend to have the stress early (in English, especially nouns) – but not all	→ 15.2

16 ←	... lack internal agreement as opposed to phrases
17 ←	... have an important subtype: verbal compounds

18 ←	Compounding and other processes: compounds may be inputs to derivation
19 ←	The meaning of compounds: transparent or opaque? Compositional or unpredictable?
20 ←	jack-in-the-box, brother-in-law, attorney general – are these compounds?
21 ←	A derivation—compound—phrase gradience

22 ←	Compounds are non-separable — all are, since they are <i>words</i>
23 ←	Compounds are referential atoms — all are, since they are <i>words</i>

1.1

Lexicalization

Lexicalization (of a compound, or any other expression) is understood as its *memorization*, i.e. *storing* in memory (= the *mental lexicon*), and *retrieving* from memory whenever need arises.

Mousepad, paper clip, egghead, dishwasher, book launch, railway, hard disk, coffee shop, blue-green, stir-fry, within are obviously all *lexicalized*, i.e. *listemes*, i.e. listed, i.e. items in the “list” of the mental lexicon. They are not put together “on the fly”, i.e. online from their components, but used as ready-made wholes.

Phrasal verbs – especially the idiomatic ones (e.g. **come to** ‘regain consciousness’ or **put up** ‘tolerate’ or ‘provide accommodation’ as opposed to **come in**), and especially the more frequent ones, are lexicalized. They are *not* compounds because they are not inseparable, i.e. not words (see 22. below).

The extreme end of lexicalization is fossilization, where the members of the original compound are hardly visible any more: **lord, lady, and nostril** used to be compounds – are compounds diachronically. See 19. below.

1.2

Whether the compounds **language description**, **Alsace-Lorraine**, **morphology textbook**, **summer day** are lexicalized depends on the speaker. On the other hand, **morphology textbook cover**, **language description file**, **coffee shop assistant**, **computer hard disk repair**, **second language acquisition research** are compounds but they are probably not lexicalized, and are produced “online” every time. The line, of course, is difficult to draw. If an expression is frequent enough (for a speaker, for a group of people, or in a given profession), it will be lexicalized.

2.1

Most compounds in English (and in other languages) indeed happen to be nouns (see the random list in 1.1). The largest subclass of N compounds consists of NN (**mouse pad**, **peace treaty**), and less commonly, AN (**madman**, **hard disk**). Second place goes to adjectival compounds (**blue-green**, **leaf green**).

2.2

There are numerous examples of non-N compounds. Let us look at some of our data – English examples first.

Notice that:

- The category of the compound where it is not N is marked in **red**.
- Wherever the *right-hand* member is X, the compound itself is also X; this is true both for Ns and non-Ns.
- Those compounds that are nouns but do not have a N on the right are **green**.

①	NN teapot	②	VV stir-fry V	③	AA white-hot A	④	PP within P
	NV globetrot V		VN swearword		AN high school		PN outhouse
	NA headstrong A		VA speakeasy ¹		AV dry-clean V		PV overthrow V
	NP pickup		VP crackdown		AP blackout		PA underripe A

3.1

Compounds are predominantly made up of *major* lexical categories: N, V, A, P. (Recall that Ps are major – thematic – categories). Combinations of prepositions (**within**, **onto**) are rare, and not productive (you cannot produce new ones at will).

3.2

Very rarely, compounds contain *minor* categories; e.g. **yes-man** ‘person (esp a man) who agrees with everything someone says’; **me-first** ‘egoistic’. Something can be a **must-have**, or a **must-see**. German **ich-sucht** ‘egomania’ (lit. ‘I-mania’).²

4.1

Most compounds in English are composed of two *free forms*, i.e. *free morphemes*, i.e. independent lexemes (that can also appear on their own). All the examples in 2.2 above illustrate this.

¹ A nightclub selling alcoholic beverages during Prohibition in the USA.

² The (jocular?) feminist creation **herstory** (in so far as it exists) on the analogy of **history** belongs here.

4.2

We will see two kinds of counterexamples to this:

- (a) Compounds formed from *cranberry morphs* – the **cran-** of *cranberry* is the stock example of cranberry morph – have one free and one bound form, a bound root. While **berry** is a free form, **cran-** is bound (it can only occur in this compound). Other examples include *boysenberry*, *huckleberry*, whose first members do not exist independently.
- (b) There is a special kind of complex expression – the *combining form* – that may be considered a compound. The expressions **biology**, **geometry**, **photography** are combining forms. The process of combining them is sometimes termed *neo-classical compounding*.

There are *initial* combining forms (first members: **astro-**, **biblio-**, **bio**, **geo-** and **xeno-**), and *final* combining forms (which can be final members: **-naut**, **-graphy**, **-logy**, **-phobia**). Some can be both initial and final (**morph-** and **-morph** as in *morphology* and *polymorph*; **phil-** and **-phile** as in *philosophy* and *Anglophile*). These combining forms – Latin and Greek roots – are obviously not affixes, since affixes cannot combine with each other.

Some linguists do not consider these as compounds exactly because not lexemes are involved in them.

Note that while cranberry words have *one* bound form, these neoclassical compounds have *two bound* members.

5.1

Ideally, there are no inflections (of any kind) within a compound. (This basically means no plural or case with nouns, and no tense or person marking in the case of verbs).³

Even if a **horse thief** steals many horses, not just one, s/he is not called a ***horses thief**. Oddly, even when a noun is usually used in the plural, such as **trousers**, it appears in the singular when inside a compound: **trouser press** and not ***trousers press**; there is no ***trousers clip** (= **bicycle clip**) or ***trousers pocket**, just **trouser clip** and **trouser pocket**.

Mousetrap catches not one mouse but more, yet it is not called ***micetrapp**.

While the **swearword** type is possible, ***sworeword** or ***swearsword** are not – no verb inflection is allowed on the non-head member.

5.2

There are, however, exceptionally, quite a few (and an apparently growing number of) English compounds whose left member, the non-head, *is* pluralized. Examples include **overseas investor**, **parks commissioner**, **programs coordinator**, **arms-conscious**, **sales-oriented**, **pants-loving**, **teeth marks**, **prisons inspectorate**, **systems analyst**, **incomes tax**, **fares table**, **sports girl**. These include both regulars and irregulars.

In that light, the Turkish **el-i-açık**⁴ (literally ‘hand-POSS-open’), which means ‘generous’, is very odd. Here, not two lexemes but an inflected word form and a lexeme have been compounded.

6.1

³ There is a silly linguistic joke in Hungarian where **szakállamtitkár** is analyzed as if it were **szakáll-am-titkár** ‘my-beard secretary’ (**szakáll** is beard; the **-(a)m** is a possessive suffix corresponding to E. **my**). The real structure, of course, is **szak-állam-titkár**, i.e. ‘special state secretary’. For the element **szak-** see the German **Fach-staat-s-sekretär**).

The trouble is that such a word cannot exist in Hungarian. No compound with such a structure is allowed because compounds cannot have *internal inflections* (i.e. inflections on the non-head) such as the **-am** here.

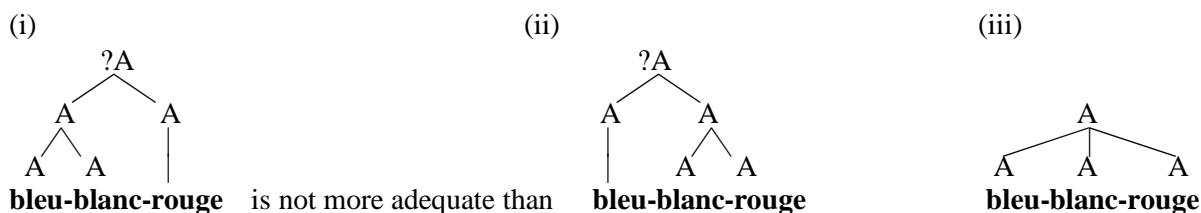
⁴ If Hungarian had a lexeme like this, it would be **?kezetárt** ‘hand-POSS-open’.

Typical compounds are *binary*, i.e. consist of two members. Most lexicalized compounds have just two members. Compounds may, however, have more branchings, i.e. more layers – but these, too, have a binary structure. This will be taken up in 7.1.

Coordinative compounds (13.2) are no exceptions to this: they are mostly binary.

6.2

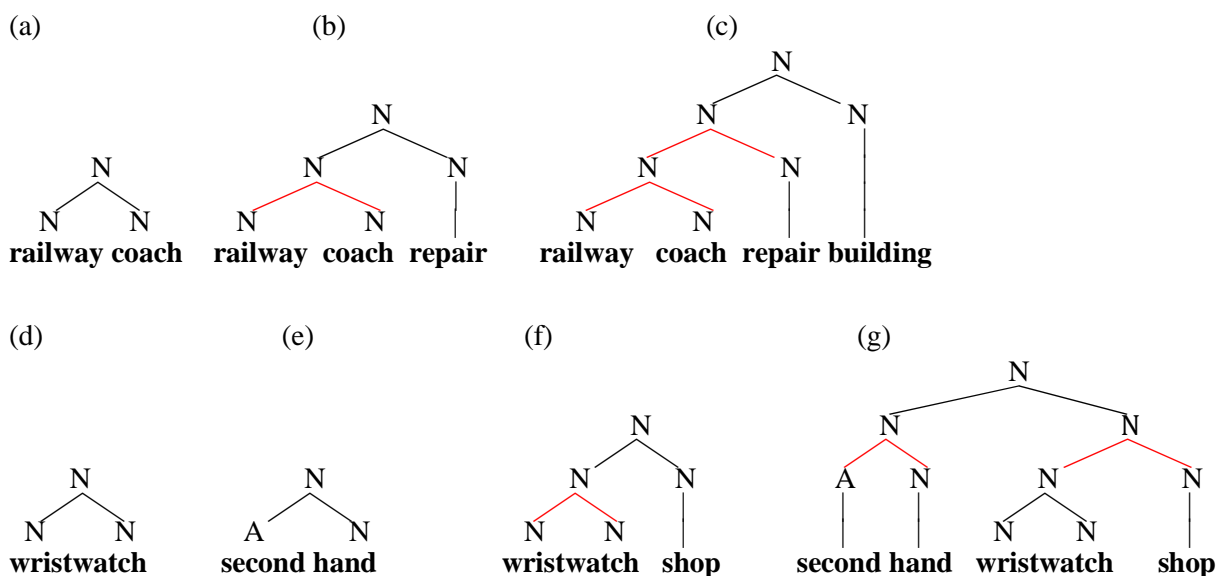
Multiple membership, i.e. *ternary* (triple) branching, however, is also possible when real-world concepts require it, e.g. in compounds denoting flags. Cf. the French compound **bleu-blanc-rouge** ‘blue-white-red’ (or the same E. word, **blue-white-red**) for the French tricolour. The name of the company **Cadbury Schweppes Hudson** (in New Zealand) is also ternary. *No two* of the three members belong together more tightly: (i) below is no better and no worse than (ii). Then probably (iii) is the adequate analysis.



Compounds with ternary branching – which, it may be claimed, have no structure at all – are very rare.

7.1

Typical compounds are binary even those that have several layers, i.e. branchings. The structure may be consistently left-branching – as in (b), (c) and (f) – or have branching on both their left and the right – as in (g).



7.2

Obviously, if (rarely) a pattern is not binary (but ternary), it cannot be binary on higher levels, because there will not be more levels, since it will have no structure in the first place: **Budapest–Paris–London** [trip].

If it is binary but coordinative (13.2), there simply will not be more levels either.

8.1

In prototypical (and most English) compounds, there is nothing between the two members (lexemes). The following Hungarian compounds also illustrate this. The English compounds given as equivalents behave identically, and they could also be used as examples here:

szem ‘eye’ + **orvos** ‘doctor’ > **szemorvos** ‘eye doctor’

év ‘year’ + **szak** ‘stage’ > **évszak** ‘season’

munka ‘work’ + **ruha** ‘clothes’ > **munkaruha** ‘work clothes’

8.2

There may, however, be linking elements – *interfixes*, or *intermorphs* – between the two members a compound:

E. **gasometer**; in **sportscar** the **-s** is not plural.

Cf. German

Auge ‘eye’ + **Arzt** ‘doctor’ > **Augenarzt** – cf. the Plural of **Auge**: **Auge-n**

Frau ‘woman’ + **Arzt** ‘doctor’ > **Frau-en-arzt** – cf. the Plural of **Frau**: **Frau-en**

Jahr ‘year’ + **Zeit** ‘stage, time’ > **Jahreszeit** ‘season’ – cf. the Possessive of **Jahr**: **Jahr-es**

Land ‘country’ + **Kunde** ‘lore’ > **Landeskunde** – cf. the Possessive of **Land**: **Land-es**

Arbeit ‘work’ + **Anzug** ‘clothes’ > **Arbeitsanzug**

This **-s-** must be a “real” interfix: **Arbeit** is feminine, and this **-s** is not a feminine ending (this form is not shared with anything else).

Recall also the German **Fachstaatssekretär**, which shows such an interfix **-s-**.

Cf. also Polish **krajoznawstwo** ‘study of national customs’ (literally: ‘country lore’; cf. **Landeskunde**), and Russian **stranovedeniye** ‘country lore’.⁵

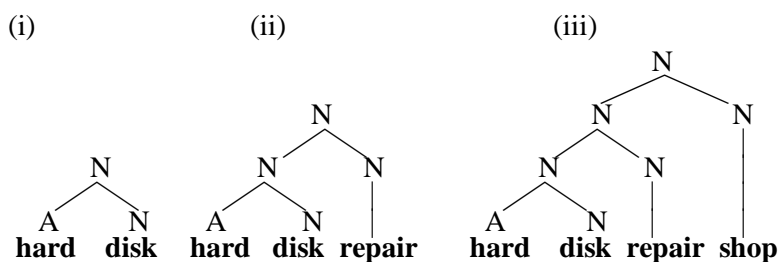
Other Russian examples include **vodopad** ‘waterfall’, **samoletostrojenije** ‘aircraft construction’, **tovaroborot** ‘merchandise circulation’, **listopad** ‘falling of leaves’.

9.1

Consider the following compounds: **hard disk**, **hard disk repair**, **hard disk repair shop**.

Their tree structures show that the rule “a N can be composed of X and a N (where X may itself be a N) can be repeated (reiterated) all over again; this is *recursion*”:

(i) **hard disk** is $N \rightarrow AN$; (ii) **hard disk repair** is $N \rightarrow NN$; (iii) **hard disk repair shop** is also $N \rightarrow NN^6$



⁵ **Landeskunde**, **krajoznawstwo** and **stranovedeniye** all mean ‘study of the geography, history, and civilization of a country’. Cf. H. **országismeret**.

⁶ 7.1 (c) and (g) are also good examples.

The German language is notorious for its freedom of recursive compounding. Multiple-layered compounds include some rather long but quite normal ones such as

Minder-wertig-keit-s-gefühl ‘feeling of inferiority’ [literally: ‘inferiority feeling’] or

Unabhängigkeit-s-erklärung ‘declaration of independence’ [lit. ‘independence declaration’]

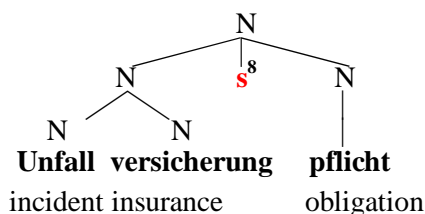
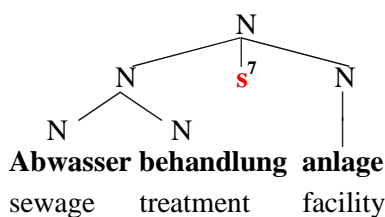
and (admittedly rare) monstrosities such as

Rind-fleisch-etikettierung-s-überwachung-s-aufgabe-n-übertragung-s-gesetz

‘beef labelling supervision duties delegation law’

(with **-s-** interfixes and an **-n-** interfix between members).

Please note, however, that e.g. the nouns **Abwasser-behandlung-s-anlage** ‘sewage treatment facility’ and **Unfall-versicherung-s-pflicht** ‘incident insurance obligation’ (with **-s-** interfixes), which have three members, are not “worse” (though they look “weird” because of the solid – one-word – German spelling) than the English **hard disk repair shop**, which has four members:



9.2

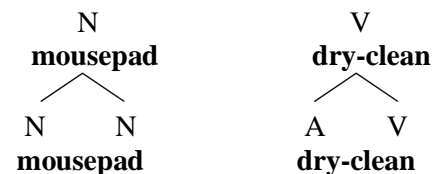
Coordinate and ternary compounds obviously are not like this. They are not recursive.

10.1

The head of a compound

The compound **sea bird** is a kind of bird; **mousepad** is a pad; a **house mouse** is a type of mouse; if you **dry-clean**, you **clean**; if you **type-write**, you **write**. Note that (what is denoted by) the entire compound *is a kind of* (the thing denoted by) the head.

The sense of the compounds in here is determined by one of the members. This is the *semantic head*. Compounds that have a head are *endocentric*.

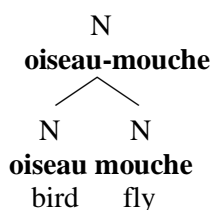


The entire compound is a *hyponym* (= *subordinate*) of the head, which is the *hyperonym* (= *superordinate*).

⁷ Recall that this cannot be a plural or genitive either.

⁸ Cannot be either plural or genitive.

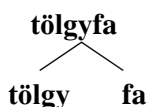
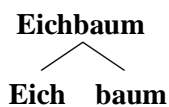
The French **oiseau-mouche** ‘hummingbird’ (literally: ‘bird-fly’), which has the head on the left, is a hyponym of **oiseau**, since a humming bird is a kind of bird.



The hyponymy–hyperonymy (subordinate–superordinate) relation is even more special in *pleonastic* compounds (= tautological) such as **oak tree** (also: German **Eichbaum** ‘oak tree’, Hungarian **tölgyfa** ‘oak tree’).

To put it simply: **oak tree** = **oak**.

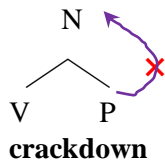
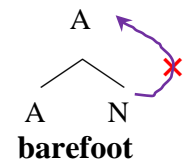
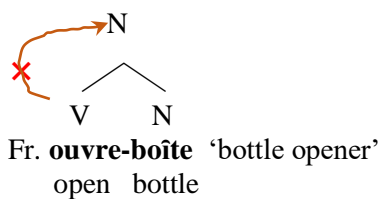
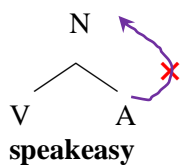
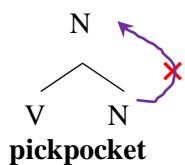
In these pleonastic compounds, not just the entire compound (**oak tree**) is a subordinate of the head (**tree**); the non-head (**oak**) is also a subordinate of the head (**tree**).



10.2

There are compounds whose sense is not determined by either member. These are *semantically headless*.

Pickpocket means a person, not a pocket; **speakeasy** means a place; **barefoot** has an adjectival meaning in **She was barefoot** and an adverbial one in **She went barefoot**; **crackdown** means severe measures (e.g. on crime). All this semantic information cannot come from either member – these compounds have no semantic head.



There is a special type of compound which is like this: **bigmouth**, **redcoat** and **egghead** are examples. A **bigmouth** is a person who is indiscreet; an **egghead** is an intellectual; a **redcoat** is (was) a British soldier. These have no semantic head, but *metonymically* (i.e. on a part-for-the whole basis) refer to people: a **bigmouth** is a person who *has* a big mouth (*metaphorically*); a **redcoat** is a soldier who *has* a red coat; an **egghead** is a person who *has* an egg-shaped head. This is why they are called possessive, or “has-a” compounds.

Note that both metonymy and metaphor have been used in the explanation above.

11.1

Typical compounds are also *endocentric* not just semantically but structurally, i.e. they have a *head* in the *grammatical* sense.

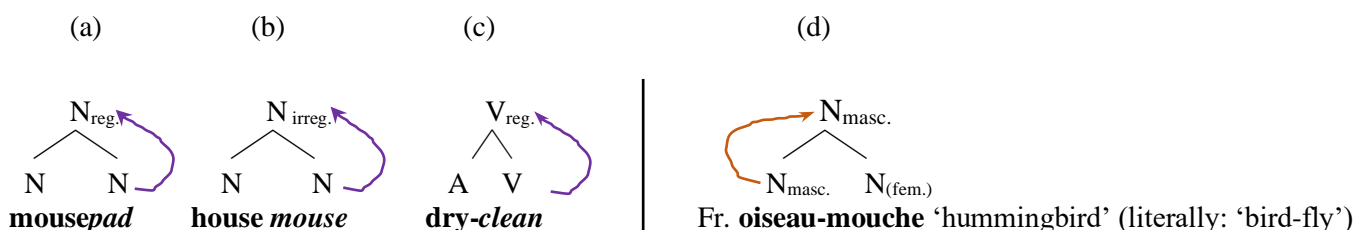
dry-clean is a regular V. This is determined by the 2nd member, **clean**, a regular V.

type-write is an irregular V. This is determined by the 2nd member, **write**, an irregular V.

mouse pad is a N and it's regular because the second member, **pad**, is a regular N.

house mouse is a N and it's irregular because the second member, **mouse**, is an irregular N.

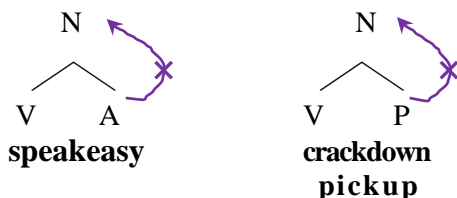
The morpho-syntactic features (e.g. word class and regularity) *percolate* or “filter” up to the compound word’s level, so the compound *inherits* them from the head. This happens in (d) too, where the head is on the left.



In languages with gender, the compound’s *gender* is determined by the head’s. Consider these German examples:
Handschuh ‘glove’ is masculine, because the head, **der Schuh**, is masculine (while **die Hand** is feminine).
 neuter **das Brot** → **das Butterbrot** ‘bread and butter’ (while **die Butter** is feminine)
 masculine **der Tag** → **der Geburt-s-tag** ‘birthday’ (while **die Geburt** is feminine)

11.2

There are compounds, however, whose category is not identical to, i.e. not determined by *either* member. The compounds **speakeasy**, **crackdown**, **pickup** show this. **Speakeasy** is N with VA structure; **crackdown** is a N with VP structure.



These compounds have no structural head: they are *headless*. We cannot explain how they are assigned a category, if not from the head. Nor can we explain how they get the rest of their grammatical features.

Once **speakeasy** is a noun, it is expected that it should have a (regular) plural: **speakeasies**. The plural feature, however, obviously cannot come from the adjective **easy**. The same goes for **crackdown** and **pickup**, where a preposition cannot be pluralized, yet these do have a plural. Pluralization obviously happens *after* compounding.

(A) There is a type of headless compound whose left-hand member is a verb and whose right-hand member is an *argument* of this verb: **pickpocket** and **ceasefire** are well-known illustrations. It is **Taschendieb** ‘pocket thief’ in German, and **zsebtolvaj** ‘pocket thief’ in Hungarian (visibly from German), and in English it could indeed be **?pocket picker**, i.e. someone who picks pockets.⁹

Crucially, these compounds are grammatically headless: the plural of the compound **carry-all** is **carry-alls**, and that of **cutwater** is **cutwaters** – but these forms clearly do not come from the right-hand member (**all** has no plural!), which is thus not the head.

(B) The special type of headless compound, which was mentioned in 10.2 (**bigmouth**, **redcoat**), is illustrated here by **bigmouth**, **sabretooth** (a kind of tiger), and **egghead**. Recall that we called them possessive compounds – “has-a” compounds – because they denote beings that have big mouths, sabre teeth, or egg-shaped heads.¹⁰ Also recall that they metonymically refer to these beings. (By contrast, remember that headed compounds are not “has-a” but “is-a” compounds: they denote something that *is a* kind of the head: **mousepad** is a kind of pad, **hard disk repair shop** is a kind of **shop**).¹¹

These compounds seem to have a *grammatical* head – on the right. The plural of **redcoat** is regular just like that of **coat**.

Sometimes, however, an odd thing happens: these grammatically headed but semantically headless compounds have a *deviant* plural – i.e. this time, a regular one instead of the expected irregular. When **sabretooth** means the tiger (not the tooth), i.e. it is semantically headless, its plural is usually **sabretooths**. The plural of **flatfoot** [slang for] ‘police officer’, another semantically headless compound, can be **flatfoots**; that of **still life** may be **still lifes**; the plural of **lowlife** is **lowlifes**.

Why should that be? Both the semantic features and the grammatical features of compounds are supposed to come – percolate up – from the head. When, however, the compound is semantically headless, this head-to-compound “information pipeline” is (partly) shut down, or **blocked**. Because of that, grammatical information can also not pass through, i.e. cannot be inherited by the entire compound:



⁹ Examples include **turnkey**, **drawbridge**, **spoilsport**, **spitfire**, **cutwater**, **turncoat**, **carry-all**, **scarecrow**, **daredevil**, **ceasefire**, **stopgap**, (John) **Lackland**, (William) **Shakespeare**.

¹⁰ Further examples include **bluestocking**, **dimwit**, **babyface**, **bluebeard**, **yellowbelly** ‘a coward’, **humpback**, **bluenose**, **redneck**.

¹¹ Because of this, the compound **jailbird**, which denotes a person (metaphorically: an inmate), is not like **bigmouth** and **egghead**, so it is not headless; it is not a person who *has a* bird.

12.1

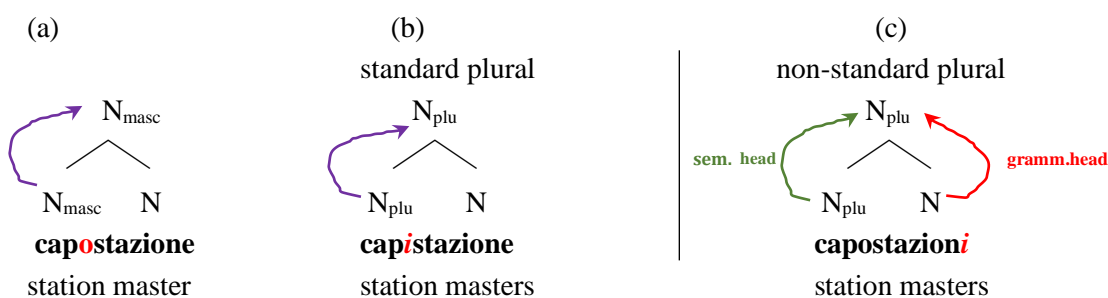
In most languages, compounds have a tendency to be right-headed, i.e. to have the head member on the right, and the non-head (the modifier) on the left. The compounds **station master** and Hungarian **állomásfőnök** ‘station master’ (lit. ‘station chief’) have their head on the right: their heads are nouns, so they are nouns; and the compounds are hyponyms of the compounds themselves (a **station master** is a kind of **master**).

Most headed compounds so far have been like that: right-headed.

12.2

A minority of languages (can) have left heads in their compounds. This relative position of the head (since headedness is involved) is also meant both *structurally* and *semantically*.

The Italian **capostazione** literally ‘head-station’ (‘station master’) illustrates this. A diagram shows what is going on here: not only is **capostazione** a kind of **capo**, the plural of **capostazione** is formed on the (left-hand) head: **capistazione**. The gender of **capostazione** is inherited from the gender (masculine) of **capo**. (**Stazione** is feminine) – see (a) and (b).



Because, however, Italian plurals attach to the right, as a rule, there is a more uniform – but non-standard – plural form for this word: **capostazioni**. The non-standard variant is more uniform, thus more systematic if you like.

However, this non-standard variant – (c) – is non-uniform from the point of view of the head properties, i.e. there is a clash between structural headedness and semantic headedness: (a) semantically, the head is still **capo**; (b) grammatically, the inflectional head is on the right side.

Note that, as is very often the case, the nonstandard form is more “logical”, or “uniform” than the standard form.

Similar compounds include **portabagagli** ‘baggage carrier’ (lit. ‘carry-baggage’); **stuzzicadenti** ‘toothpick’ (lit. ‘pick-teeth’); **portacenere** ‘ashtray’ (lit. ‘carry ashes’), **portacappelli** ‘hat box’ (‘lit. carry hats’), **portalettere** ‘postman’ (lit. ‘carry letters’), **grattacielo** ‘skyscraper’ (lit. ‘scrape sky’); and even **portastuzzicadenti** ‘tooth pick carrier’ (literally: ‘carry pick teeth’).

Cf. also Fr. French **grille-pain** ‘toaster’ (literally: ‘toast bread’), or the Fr. **ouvre-boîte** ‘bottle opener’ (literally: ‘open-bottle’) above.

13.1

Most compounds (in all languages) are *subordinate*, whether headed or not; whether right- or left-headed, etc. One of the members – the non-head – is both semantically and structurally subordinate to the other. In the German **Taschendieb** ‘pickpocket’, (literally ‘pocket thief’), the non-head **Tasche** is subordinated to the head **Dieb**.

This holds even between members of exocentric compounds: in **pickpocket**, neither member is head, but **pocket** is still subordinated to **pick** (it is the argument of **pick**). The modifier **loud** is subordinated to **mouth** in **loudmouth**. Although the primary semantic head designating a person who picks pockets is not expressed, **pocket** is – secondarily – the morphological head that determines inflection.

Subordinate compounds appear to be always binary (though their structuring may be recursive).

13.2

The dispreferred alternative type – *coordinative* compounds – have members of equal status, cf. the nouns **speaker-hearer**, **washer-dryer**, **actor-director**; the adjectives **bittersweet**, **English-French** [e.g. dictionary]. In nouns, the plural attaches to the right member or both members: **speaker-hearers** but **women doctors**. The verb **stir-fry** can be argued to be coordinate (you stir *and* fry) or subordinate (this is a kind of frying).

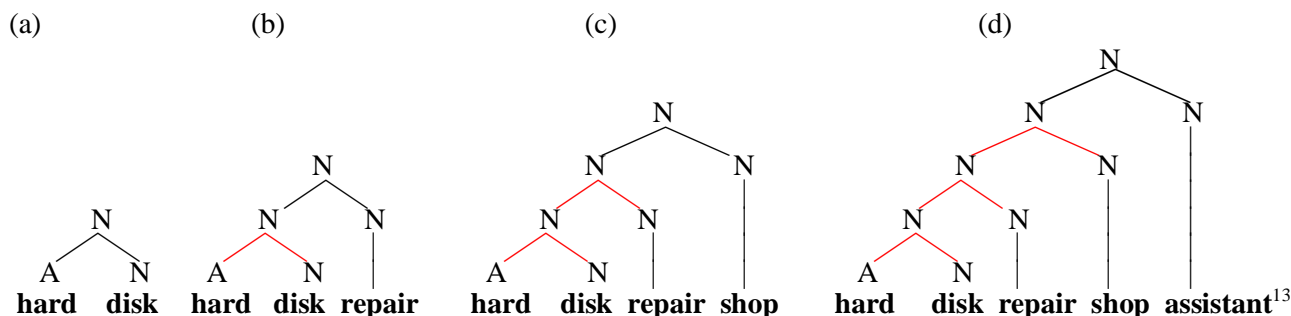
There are different semantic types of coordinate compound depending on the internal relation of the members: **speaker-hearer** or **bitter-sweet** are *appositional*. If something is **bittersweet**, it is both **bitter** and **sweet** at the same time. **Moscow-Washington** [e.g. flight] is not appositional; **Alsace-Lorraine** (G. **Elsaß-Lothringen**) is the combination of these two things; **English-French** is again different (probably more similar to **Moscow-Washington**). Is **Austria-Hungary** like **Alsace-Lorraine**? Is it one entity, or a pair?

Parent-teacher [association], **mother-daughter** [relationship] or **patient-doctor** [relationship] are all similar, but they are unlike the other examples here.¹²

The Italian **caffè-latte** ‘coffee with milk’ can be argued to be coordinate; because, however, its order cannot be reversed, it can also be treated as (left-headed) subordinate. Cf. the right-headed German **Milchkaffee** ‘coffee with milk’: the (potential) inverted **Kaffeemilch** would mean ‘milk with (a drop of) coffee’.

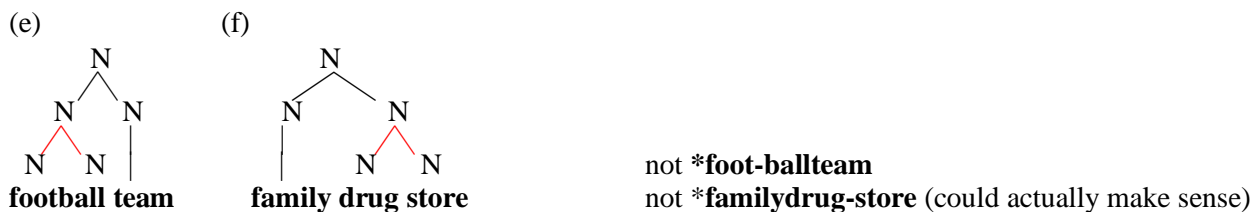
14.1

Consider the compounds (a)–(d). Their analysis is as follows; the tree structures are left-branching:



14.2

Now compare the hierarchical structure of the nouns **football team** and **family drugstore**.



(e) is **left-branching**, like (b)–(d) above.

(f) is **right-branching**. Right-branching is less frequent in English.

Now try your hand at **ball point pen** and computer **hard disk**. Are these left or right-branching?

¹² Properties of coordinate compounds differ from language to language: the order of their members is determined not grammatically (since the members are equal), but (i) pragmatically: e.g. more important first (**speaker-hearer**); (ii) prosodically: e.g. longest last.

¹³ You might try your hand at **hard disk repair shop assistant training course**.

15.1

In languages where this applies, the **compound stress rule** (primarily for nouns) is different from the stress rule for phrases. The compound rule assigns a single stress to the first member: the stock examples again are

'blackboard and 'blackbird (compounds) ↔ ,black 'board and ,black 'bird (phrases)
Hung. 'drágakő 'precious stone' (compound) ↔ ,drága 'kő 'expensive stone' (phrase)

15.2

Unfortunately, by far not all compounds behave like these, even for nouns. There are different kinds of deviations: unique exceptions, differences by word class to word class, differences by individual speakers or groups of speakers.

16

Internal agreement

In languages where this applies, compounds may differ from phrases because phrases show **internal agreement**.

The German compound **Rotkohl** means 'red cabbage' (i.e. a special type of cabbage). In the noun phrase **roter Kohl** '(any) red cabbage' the adjective has an ending **-er**: an agreement inflection. The adjective must agree with the noun for gender, number, and definiteness (**Kohl** is singular/masculine). In the compound, the adjective **Rot** is not inflected.

The German **Maul** is masculine, e.g. **ein großer Maul** 'a big mouth', but in a compound it is **Großmaul** 'loudmouth'.

The Dutch compound **grot-hoek lens** means 'wide-angle lens'. **Groot** means 'wide'. When **groot** is used with the feminine noun **hoek**, the form **grote** must be used: **grote hoek**.

compound

phrase

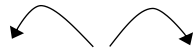
 N
 / \
 A N
Rotkohl
Großmaul
groot-hoek

 NP
 / \
 A N
roter Kohl
großer Maul
grote hoek

There is an easily identifiable and very important – productive – type of compound: the *verbal* or *synthetic* or *secondary* compound. Unfortunately, “verbal”, “synthetic” and “secondary” as terms do more harm than good. “Verbal” may suggest that they are verbs: they never are. “Synthetic” and “secondary” are not helpful either.

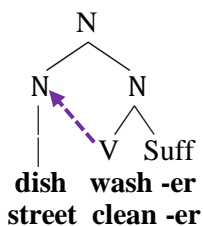
Most compounds seen so far are not verbal: they are *root* or *primary* compounds. (Another pair of not-too-useful terms).

A verbal compound has a verb at its *base* and a modifying element that could function as argument of this verb.



Recall e.g. that **bake** takes 2 arguments: **Susan baked a cake**

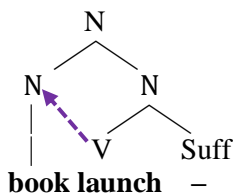
The compound **dishwasher** has **wash** at its base, and the first member, **dish**, is its argument: **wash dishes**.



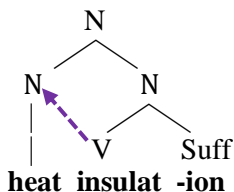
Street cleaner has **clean** at its base, and **street** is its argument: ‘clean (a/the) street’. **Cake baker**, another verbal compound, has **bake** at its base, and **cake** is its argument.

Language description is somewhat trickier: it has a variant of the root **describe** (= **descript-**) at its base, and the first member, **language**, can be seen as its argument (describe a language).

Book launch is again special: it has the verb **launch** at its base, which is converted into a N (**launch**) without any affix (this is the essence of *conversion*), and the first member, **book**, can be seen as its argument: **launch books**.

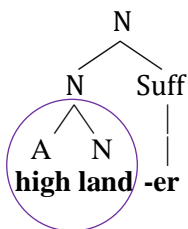


Thus, in most verbal compounds the non-head (the modifier) is some kind of argument, most frequently – but not necessarily – an object. **Heat insulation** is not insulating the heat but insulating against it; still, **heat** is the argument of **insulate**:

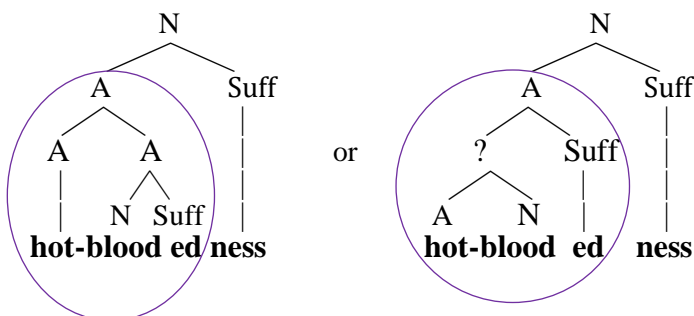


18

Compounds may be further derived, and in such cases the products are derivations, not compounds: **highlander** is derived from a compound, but that compound is a N, and by the time of the derivation, its internal structure is no longer relevant.



Hot-bloodedness is derived from a compound, but that compound is an A, and by the time of the derivation, its internal structure is no longer relevant.



19

Compounds: transparent or opaque?

A root (i.e. not verbal compound) e.g. of the [N NN] type, i.e. a noun consisting of two nouns, is always potentially *ambiguous*, or at least *vague/indeterminate*, i.e. opaque and not transparent.

Consider the compounds **mosquito net**, **hair net** and **butterfly net**.

There is no straightforwardly predictable semantic relationship between the two elements such compounds, so their meaning is not predictable. **Mosquito net** is for keeping mosquitos away, **hair net** for keeping hair together, and **butterfly net** for catching butterflies. These meanings, however, do not follow either from the structure of these compounds or the meanings of **mosquito**, **hair**, **butterfly** and **net** individually. If one encounters one of these for the first time, the best one can do is guess its meaning on the basis of *pragmatic* considerations, or *analogy* with other similar compounds. An **iron pipe** is *made of* iron, while a **stove pipe** is the pipe *of* a stove.

The compound **leg man** may mean e.g. a (i) worker in furniture factory working with legs (ii) pirate with wooden leg (iii) person preferring drumsticks (iv) male casting director choosing dancers by their legs.

In that sense, compounds have an *idiomatic* meaning, or maybe are *idioms* – complex expressions whose overall meaning cannot be calculated from the meanings of their parts.¹⁴

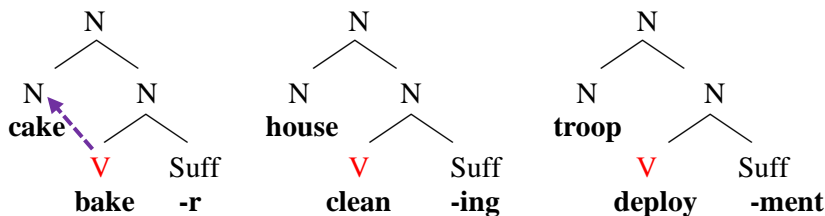
A compound may also be ambiguous between a root compound interpretation and a verbal compound interpretation. Suppose you don't know **street seller**; if you don't know even whether it is root or verbal, you will not know whether it is a person selling streets (verbal compound), or a person selling *in* the streets (root compound). The prevailing interpretation of **-er** verbal compounds is that they are nominalizations of verb phrases: **street sweeper** = 'smb/smth that sweeps streets'.

¹⁴ Kinds of *transparency* may be distinguished: transparency of both members: **DOOR BELL**; transparency of head, opacity of non-head: **strawBERRY**; transparency of non-head, opacity of non-head: **JAILbird**; opacity of both members: **humbug**.

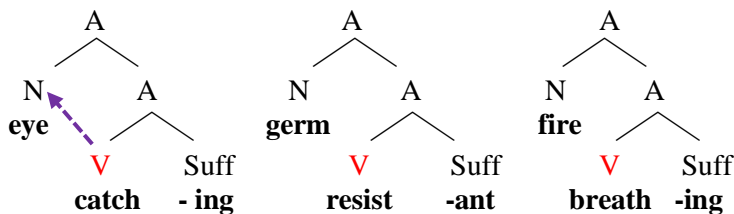
When compounds are lexicalized, they may reach a stage – the extreme end of lexicalization – where they are **fossilized**: the members of the compound are not (or hardly) visible: in OE. **lord** used to be analyzed as **hlāfweard** (lit. ‘loaf keeper’); **lady** was a compound, **hlāfdige**, meaning ‘loaf kneader’.¹⁵

By contrast – as we have seen – the interpretation of verbal compounds is very specific.

Nouns



Adjectives



20

Are these compounds?

Some authors include the following types among compounds: (a) **jack-in-the-box** (b) **brother-in-law** (c) **attorney general**.¹⁶

(a) The **jack-in-the-box** looks like an NP [N-P-D-N], where the head is modified by a PP (such e.g. as **noise in the street**). These, however, *standardly* form their plural not on their heads but on the *whole*: [**jack-in-the-box**]**es**. This is obviously not possible with real phrases: *[**noise in the street**]**s** is not the plural of **noise in the street**. This suggests that this type is a *word*.

Many speakers, by contrast, pluralize these items on the head: **jack**s**-in-the-box** – i.e. they treat them as *phrases*, not (compound) words.

(b) The **brother-in-law** type *standardly* forms its plural on the head, so it must be a *phrase*: **brother**s**-in-law**.

Many speakers, however, pluralize these on the whole, i.e. at the end; they treat them as (*compound*) *words*. (Not all expressions in this type are uniform: the plural of **attorney at law** can only be **attorney**s** at law**.)

(c) The **attorney-general** type can *standardly* form their plural in *two* ways: [**attorney general**]**s** or **attorney**s** general**.

The form **attorney general**s**** suggests that this is a word; **attorney**s** general**, suggests that it is a phrase.

Note that this way or other, type (c) is not a simple case: *if* it is a compound, it is left-headed; *if* it is a phrase, then the adjective “is on the wrong side”.

¹⁵ The Hungarian word **arc** ‘face’, and E. **blackguard**, **breakfast** and **curfew** are also compounds diachronically. The latter two are special. Check all of them in a dictionary.

¹⁶ Also: **johnny-on-the-spot**; **mother-in-law** (etc.), **man about town**, **editor-in-chief**, **right of way**; **court martial**, **president elect**, **heir apparent**, **poet laureate**.

21

Derivation—compound—phrase: a gradience

Compounds seem to be situated between derivations and phrases. As we have seen, one issue in the analysis of compounds is exactly their demarcation from phrasal expressions. A sequence might be either a compound (hence a word), or a phrase – and it is not easy to decide which one it is.

On the other side of compounds are derivations, and demarcation is a problem here too. The crucial distinction between compounds and derivatives is that in (prototypical & English) compounds the constituents are lexemes, whereas derivatives involve affixes.

One reason why this distinction is not clear-cut is that a lexeme may develop into a derivational morpheme. One example will suffice: **man** /mæn/ is evidently an (irregular) noun (and its plural is) /men/. When, however, it is (some of them: *was*) used in compounds such as **postman, mailman, policeman, snowman, repairman**¹⁷, it does not (simply) mean ‘male human being’, and importantly, it is often not pronounced /mæn/ but /mən/. This might show that it is on its way to becoming an affix.

22

Argument for compound-hood from wordhood

The most important feature of a compound – whether (right)headed or not; whether lexicalized or not etc. – is that it is *inseparable*, since it is one word: no material may come between the members of a compound.

This is how syntactic phrases like **high school** ‘tall school building’ differ from compounds such as **high school** ‘secondary school’. An adjective can be inserted in the phrase, but not in the compound: **high, big school** (phrase) vs. ***high big school** (compound).

Adjectives can, of course, modify the compound as a whole: **big [high school]**.

Phrasal verbs may be lexicalized (may be listemes), but are not compounds because they are not inseparable: **come to** ‘regain consciousness’ may be separated by adverbs: **come quickly to**.

Phrasal verbs – especially the idiomatic ones (e.g. **come to** ‘regain consciousness’ or **put up** ‘tolerate’ or ‘provide accommodation’ as opposed to **come in**) and especially the more frequent ones, are lexicalized.

This is also true of Hungarian *prefixed verbs*, which are not compounds – are not words – by this criterion; the verb in them can be separated from the prefix by different things:

letol ‘tell off’ – but **le is tol** ‘tell off too’ and **nem tol le** ‘not tell off’, etc.

Similarly in German verbs:

aus-gehen ‘to go out’ but

geht...aus in 3sgPres (e.g. **sie geht schnell aus** ‘she goes quickly out’),

ging...aus in the past (e.g. **sie ging schnell aus** ‘she went quickly out’)

aus-ge-gang-en in the past participle (**sie ist schnell ausgegangen** (‘she went quickly out’).

23

The structure of compounds is not accessible to the syntax: it is as if once a compound has been formed, it is as if it is “closed down” by a pair of brackets. Once **truck driver** has been combined from its constituents, the non-head cannot usually be referred to anaphorically (i.e. backwards); compounds are *anaphoric islands*. It is impossible to say ***Truck drivers never fill them up completely** because you cannot refer back to **truck**.

Once **mousetrap** has been combined from its members, the non-head is “lost” for the syntax. Similarly to the **truck driver** example, ***She put out mousetraps because she’s afraid of them** is unacceptable to most speakers, because **them** cannot refer to **mouse**.

¹⁷ Also: **layman, anchorman, chairman, spokesman, foreman, Frenchman, Chinaman** (etc.), **repairman**

Most titles between “ and “are chapters (sometimes accompanied by a number); some are articles in collections.

- Adams 2013. “Noun compounds” (6); “Adjective compounds” (7); “Verb compounds” (8), in Valerie Adams *Complex words in English*, Routledge.
- Baker & Hengeveld 2012. Anne E. Baker & Kees Hengeveld eds. “Compounds and Idiomatic Expressions”, in *Linguistics*, Blackwell.
- Bauer 1998. Laurie Bauer “When is a sequence of two nouns a compound in English?” in *English Language and Linguistics*, 2, pp 65-86.
- Bauer 2003. Laurie Bauer “Processes involving several lexemes”, in *Introducing linguistic morphology* 2nd ed. Edinburgh University Press.
- Bauer 2005. Laurie Bauer “The borderline between derivation and compounding”, in Wolfgang U. Dressler et al. eds. *Morphology and its demarcations*, Benjamins.
- Bauer 2006a. Laurie Bauer “Compounds and minor word-formation types”, in Bas Aarts & April McMahon eds. *The handbook of English linguistics*, Blackwell.
- Bauer 2006b. Laurie Bauer “Compound”, in Keith Brown *Encyclopedia of language and linguistics* 2nd ed. Elsevier.
- Bauer 2009. Laurie Bauer “Typology of Compounds” (Ch 17), in Rochelle Lieber & Pavol Štekauer eds. *The Oxford handbook of compounding*. OUP.
- Bisetto & Scalise 2009. Antonietta Bisetto and Sergio Scalise “The Classification of Compounds” (Ch 3), in Rochelle Lieber & Pavol Štekauer eds. *The Oxford handbook of compounding*. OUP.
- Booij 2005a. Geert Booij “Compounding and derivation: Evidence for Construction Morphology”, in Wolfgang U. Dressler et al. eds. *Morphology and its demarcations*. Benjamins.
- Booij 2005b. Geert Booij “Compounding”, in *The grammar of words. An Introduction to Linguistic Morphology* 2nd ed. OUP.
- Brinton & Brinton 2010. “Compounds”, in Laurel J. Brinton & Donna M. Brinton *The Linguistic Structure of Modern English*. Benjamins.
- Bussmann 1996. Hadumod Bussmann “Composition (also Compounding), in *Routledge Dictionary of Language and Linguistics*.
- Carstairs-McCarthy 2002. “Compound words, blends and phrasal words”, in Andrew Carstairs-McCarthy *An Introduction to English Morphology*. Edinburgh University Press.
- Coates 1999. “Compound and complex bases”, in Richard Coates *Word structure*. Routledge.
- Fabb 1998. Nigel Fabb “Compounding”, in Andrew Spencer & Arnold M. Zwicky eds. *The Handbook of Morphology*, Blackwell.
- Giegerich 2004. Heinz J. Giegerich “Compound or phrase? English noun-plus-noun constructions and the stress criterion, in *English Language and Linguistics* 8.1: 1–24. CUP
- Giegerich 2009. Heinz Giegerich “Compounding and Lexicalism” (Ch 9), in Rochelle Lieber & Pavol Štekauer eds. *The Oxford handbook of compounding*. OUP.
- Haspelmath & Sims 2002. Martin Haspelmath & Andrea D. Sims “Compounding types (7.1)”; “Hierarchical structure in compounds” (7.2), in *Understanding morphology*. Hodder Education.
- Kiefer 2009. Ferenc Kiefer “Uralic, Finno-Ugric: Hungarian” (Ch 29), in Rochelle Lieber & Pavol Štekauer eds. *The Oxford handbook of compounding*. OUP.

- Libben & Jarema 2006. Gary Libben & Gonia Jarema eds. *The Representation and Processing of Compound Words*. OUP.
- Lieber & Štekauer 2009. Rochelle Lieber & Pavol Štekauer “Introduction: status and definition of compounding” (Ch 1), in Rochelle Lieber & Pavol Štekauer eds. *The Oxford handbook of compounding*. OUP.
- Lieber 2005. Rochelle Lieber “English word-formation processes”, in Pavol Štekauer & Rochelle Lieber eds. *Handbook of word-formation*. Springer.
- Lieber 2009a. Rochelle Lieber “Compounding”, in *Introducing morphology*. CUP.
- Lieber 2009b. Rochelle Lieber “A Lexical Semantic Approach to Compounding” (Ch 5), in Rochelle Lieber & Pavol Štekauer eds. *The Oxford handbook of compounding*. OUP.
- Lieber 2009c. Rochelle Lieber “IE, Germanic: English” (Ch 18), in Rochelle Lieber & Pavol Štekauer eds. *The Oxford handbook of compounding*. OUP.
- Matthews 2006. Peter Matthews “Compounds”, in *Morphology* 2nd ed. CUP.
- McArthur 1992. Tom McArthur “Compound-complex word” & “Compound word”, in *The Oxford Companion to the English Language*. OUP.
- Napoli 1996. “Compounding”, in Donna Jo Napoli, *Linguistics*, OUP.
- O’Grady et al. 1996. William O’Grady, Michael Dobrovolsky & Francis Katamba “Compounding”, in *Contemporary linguistics. An introduction*. Longman.
- Olsen 2000. Susan Olsen “Composition”, in Geert Booij et al. eds. *Morphology. An International Handbook on Inflection and Word-Formation*. Vol. 1, Walter de Gruyter.
- Peitsara 2006. Kirsti Peitsara “MAN-Compounds in English”, in *Selected Proceedings of the 2005 Symposium on New Approaches in English Historical Lexis (HEL-LEX)*, ed. R. W. McConchie et al., 113-122.
- Plag 2003. Ingo Plag “Compounding”, in *Word-Formation in English*. CUP.
- Plag 2006. Ingo Plag “The variability of compound stress in English: structural, semantic, and analogical factors”, in *English Language and Linguistics* 10.1: 143–172. CUP.
- Radford et al. 2009. Martin Atkinson, David Britain, Harald Clahsen, Andrew Radford & Andrew Spencer “Compounds”, in *Linguistics. An Introduction* 2nd ed. CUP.
- Scalise, Bisetto & Guevara 2005. Sergio Scalise, Antonietta Bisetto & Emiliano Guevara “Selection in compounding and derivation”, in Wolfgang U. Dressler et al. eds. *Morphology and its demarcations*. Benjamins.
- Selkirk 1982. Elisabeth O. Selkirk “Compounding”, in *The Syntax of Words*. MIT Press.
- Spencer & Zwicky 1998. Andrew Spencer & Arnold M. Zwicky *The handbook of morphology*. Blackwell.
- Štekauer 2005. Pavol Štekauer “Compounding and affixation: Any difference?”, in Wolfgang U. Dressler et al. eds. *Morphology and its demarcations*. Benjamins.
- Stockwell & Minkova 2001. Robert Stockwell & Donka Minkova “Compounding”, in *English words. History and structure*. CUP.
- ten Hacken 2000. Pius ten Hacken “Derivation and compounding”, in *Morphology. An International Handbook on Inflection and Word-Formation*. De Gruyter.