

# Compound and incorporation constructions as combinations of unexpandable roots

Martin Haspelmath

*Max Planck Institute for Evolutionary Anthropology*

draft, June 2023

This paper offers definitions of the terms *compound construction*, *compound* and *incorporation construction* that can be applied to all languages in the same way. The earlier literature has often expressed pessimism about identifying such elements across languages in an objective way, but I propose definitions that do not rely on notions such as “word” or “morphology”. I define *incorporation* as a special kind of verbal compound construction, and a compound construction as a combination of strictly adjacent roots. Rather than being “non-phrasal”, I say that compounds are defined as being “non-expandable” by modifiers. These definitions are shared-core definitions, like most other definitions of comparative concepts: They capture the core of the types of elements that have been called “compound” and “incorporation” in the earlier literature, but not necessarily every compound and every incorporation in every language. I hope that these definitions show that we do not need to wait for the definitive theory of compound and incorporation constructions before we can provide simple and clear definitions.

## 1. Defining “compound construction” and “incorporation construction”

In this paper, I propose and discuss definitions of the well-known terms *compound* and *incorporation* in general grammar. Both these terms are well-known to linguists, have long been used and are associated with rich stereotypes, but so far, no commonly accepted definitions have emerged. I propose the definitions in (1) and (2) of *compound construction* and *incorporation construction* as comparative concepts.

### (1) compound construction

A compound construction is a construction consisting of two strictly adjacent slots for roots that cannot be expanded by full nominal, adjectival, or degree modifiers.

### (2) incorporation construction

An incorporation construction is an event-denoting verb-noun compound construction in which the noun occupies an argument slot of the verb and occurs in a position where nominal patient arguments cannot occur.

Some examples of compounds from different languages are given in (3), and some examples of noun-incorporating verb forms are given in (4).

### (3) some compound forms

a. German	<i>Auto-bahn</i>	[car-way]	‘expressway’
b. French	<i>tire-bouchon</i>	[pull-cork]	‘corkscrew’
c. Chinese	<i>fēi-jī</i> (飛機)	[fly-machine]	‘airplane’
d. Mwoɬlap	<i>tit teñteñ</i>	[punch cry]	‘make (someone) cry by punching’ <sup>1</sup>

<sup>1</sup> François (2004: 112); Mwoɬlap is an Austronesian language of Vanuatu.

## (4) some noun-incorporating verb forms

- a. Mohawk                    *wak-tsitsia-ientho-on*  
1SG-flower-plant-ST  
'I planted flowers' (Mithun 2010)
- b. Biniñ Gunwok            *ba-warde-jobge-ng*  
3PL-rock-split-PST  
'they split the rock' (Evans 1996: 65)
- c. Guaraní                    *ai-po-pete*  
1SG-hand-slap  
'I slapped the hand' (Velázquez-Castillo 1996: 99)
- d. Mapudungun            *kintu-waka-le-y*  
seek-cow-PROG-IND.SG  
'he is looking for the cows' (Baker et al. 2004: 139)

The definitions proposed here have the following notable properties that will be discussed in the course of this paper:

- (i) They can be applied equally to all languages as they do not make reference to language-particular features.
- (ii) They are not based on the notion of 'word', so that a word can be defined with reference to 'compound', and they do not presuppose a distinction between morphology and syntax (see Haspelmath 2011).
- (iii) They are not prototype-based or fuzzy.
- (iv) They single out the great majority of constructions that have been called 'compound' and 'incorporation', as well as the most typical cases, but not all cases.

It should be noted that there is no claim that these concepts are particularly significant for linguistics, and especially the definition of *incorporation* is clearly so specific that there seems to be nothing natural about it. However, it is necessary to explain these terms to newcomers (e.g. in textbooks and encyclopedia articles) because they are so widely used, and it is best to provide clear definitions in order to avoid the impression that the terms necessarily correspond to natural categories of the world. Linguists often assume (sometimes implicitly) that grammar is by nature divided into morphology and syntax, but this view was inherited from a long tradition and is not a result of linguists' research. Thus, it is best not to reinforce the stereotypical view of compounds as belonging to "morphology" (as opposed to "phrases" that belong to "syntax"), and to define both compounds and phrases as types of constructions (see Haspelmath 2023 for the term *construction*).

Thus, this paper is a proposal to improve the methodology of general linguistics by offering clear definitions for two technical terms that are widely used by linguists. Previous overview articles often note that clear definitions are unavailable or difficult (e.g. Schlücker 2023: §2.2: "There is no clear or general definition of compound"), but this state of affairs is not unavoidable. The definitions proposed here will not satisfy everyone, but as no empirical claims are involved, this is not important. The central message is that we can have clear and relatively simple definitions of core technical terms of grammar, and that we do not need to wait until we have solutions to deep problems such as the purported syntax-morphology distinction.

## 2. Avoiding language-particular defining features

The definitions in (1)-(2) are intended as general concepts that can be applied to any language, and thus they do not contain concepts that are relevant only to particular languages. For example, German Compounds are defined with respect to a special stress pattern (e.g. Compound *Rótwein* ‘red wine’, contrasting with the Phrase *rôter Wéin* ‘wine which is red’),<sup>2</sup> but this criterion plays no role in many other languages. Modern Greek Compounds are defined with respect to the absence of inflectional affixes on the modifier root (Ralli 2013), but again, this criterion cannot be applied to all languages as not all languages have such inflectional affixes. Dressler (2006) puts it as follows:

“More explicit universal definitions of the intensional type are not only theory-dependent ... but also cross-linguistically never watertight – in many languages there are exceptions or fuzzy transitions to non-compounding.”  
(Dressler 2006: 24)

And a very similar formulation is chosen by Finkbeiner & Schlücker (2019: 9): “In [defining compounds] we do not aim for more than a rough approximation, as it is clear that the respective criteria are not only in part language-specific, but also a matter of controversial theoretical debate.” Lieber & Stekauer (2009: 7) discuss a number of proposed criteria for compoundhood but conclude that “These criteria might have limited utility within a particular language or group of languages, but cross-linguistically they cannot be definitive.” Similarly, Bauer (2017: 19) says about the criteria that he mentions: “Unfortunately, the criteria are difficult to apply across language types.”

In this paper, I propose defining criteria that are applicable across language types in the same way (without “difficulty”)<sup>3</sup>.

## 3. Compounds vs. compound constructions

The definitions in (1)-(2) characterize kinds of compound constructions, but we often also talk about compounds, i.e. kinds of forms. Thus, we should add a definition of *compound*:

### (5) compound

A compound is a form (consisting of two adjacent roots) that instantiates, or was created by, a compound construction.

For regularly formed compounds, the distinction between “instantiation” and earlier “creation” is not relevant. For example, English compounds like *snake poison* or *apple cake* or *cow tree* (‘tree that cows like to rub up against’; Downing 1977: 827) were created from the construction  $[N_2 - N_1]$  ‘ $N_1$  which is related to  $N_2$ ’, and they can be said to currently instantiate this construction because their meanings are compatible with the general semantic schema.

However, there are many compounds with non-compositional meanings which cannot be said to instantiate a compound construction, e.g. *soap opera* (not a kind of opera), *honeymoon* (not a kind of moon). However, it seems clear that they were created from the same construction as regular compounds, and the figurative nature of their meanings

<sup>2</sup> Note that language-particular terms are written in upper case here, following a well-known convention.

<sup>3</sup> Actually, I would say that defining technical terms is never difficult (see this blogpost: <https://dlc.hypotheses.org/2675>). What is difficult is to find a clear definition of an old term that applies to all languages in the same way and that will satisfy one’s colleagues.

is often still apparent. Excluding all these from the definition of compounds would give unintuitive results, but this is not a very important point (as unintuitive results cannot be avoided in general; this is discussed in §10 below).

One could object that the definition in (5) includes a diachronic criterion (“or was created from”) that should not have a place in definitions. This is a valid point, and it would alternatively be possible to define a compound simply as a form that instantiates a compound construction. The definition would then be much narrower, but it would still correspond largely to our intuitions.

#### 4. Compounds consist of roots (not stems or words)

It is sometimes said that compounds consist of two words (Marchand 1960: 11) or two stems (Schlücker 2023), but some authors prefer to remain vague: Thus, Olsen (2015: 364) says that compounding involves the combination of “two or more lexemes (roots, stems, or freely occurring words)”. Here I define compounds as consisting of roots, in line with definitions such as (6).

(6) “A simple way to make new lexemes is to make compounds by combining noun, verb and adjective roots.” (Gebhardt 2023: 133)

A root is defined as in (7), which basically says that a root is a morph that denotes an action, an object or a property (I had used the latter simpler definition in Haspelmath 2018: 316).

##### (7) root

A root is a contentful morph (i.e. a morph denoting an action, an object or a property) that can occur as part of a free form without another contentful morph.

Restricting compounds to root combinations means that the verbal prefixes of Indo-European languages (e.g. German *um-armen* ‘embrace’, Russian *vy-nimat* [out-take] ‘take out’) do not count as compound elements, which is in line with current usage.<sup>4</sup> It also means that combinations involving pronouns (e.g. English *him-self*) or adpositions (e.g. *on-to*) are not regarded as compounds, which again corresponds to the way the term *compound* is generally used.

It is not felicitous to define compounds as consisting of two words or two lexemes because words often include inflectional affixes, and we generally think of compound stems as not including inflectional affixes. Compounds may of course be inflected by adding further affixes, but the two compound constituent elements are roots. In languages with thoroughgoing case and number inflection, the compound members do not carry such affixes. For example, the Latin adjective-noun compound *magn-anim-(us)* ‘great-spirited’ includes the root *magn-* ‘great’, but not any of the case and number inflection that normally occurs on the adjective (nominative singular *magn-us*, accusative singular *magn-um*, dative plural *magn-is*, and so on).<sup>5</sup>

<sup>4</sup> As described by Olsen (2015: §1.1), the older tradition from the 19th and earlier 20th centuries (up to Marchand 1960) was to regard such forms as compounds rather than as prefixed formations.

<sup>5</sup> “Co-compounds” such as Russian *čaški-bljudca* [cups-plates] ‘dishes’ (see Wälchli 2015) and “appositive compounds” such as Russian *inžener-fizik* [engineer-physicist] ‘person who is both engineer and physicist’ (Ohnheiser 2019: 256) would thus not count as compounds by the current definition.

Or should we say that compounds consist of *stems* rather than roots? This would have the advantage that commonly found forms such as those in (8) would be included, where one of the elements is a derived form rather than a root (*free-dom*, *Heiz-ung*, *es-és*).

(8)	English	<i>free-dom day</i>		
	German	<i>Gas-heiz-ung</i>	[gas-heat-ing]	‘gas heating’
	Hungarian	<i>hó-es-és</i>	[snow-fall-NMLZ]	‘snowfall’

But what exactly is a “stem”? There is a good definition of the term *root* (as given in (7) above), but we do not have a generally accepted definition of *stem*. One could suggest that a stem is a combination of a root and derivational affixes, but this presupposes a definition of “derivational” as opposed to “inflectional”, and so far, there is no generally accepted way of drawing the line between these (but see Haspelmath 2024 for a recent proposal).

Finally, it should be noted that the restriction to roots means that phrases cannot be compound members. Thus, so-called “phrasal compounds” (e.g. Trips & Kornfilt 2015) of the type *chicken and egg situation* do not fall under the current definition. They are of course compound-like in languages in which they occur, but there does not seem to be a way of defining *compound* in general such a way that they can be included.<sup>6</sup>

## 5. The roots are adjacent

If two roots are combined but are not always adjacent to each other, we do not call the resulting pattern “compounding”. Thus, the German expression *statt-finden* ‘take place’ in (9) does not qualify as a compound, because when the verb is finite as in (9b), the noun-like part *statt* (‘place’) is not adjacent to the verb *findet*. (Thus, Wurzel’s

- (9) a. *Die Buchmesse kann dieses Jahr statt-finden.*  
 the book.fair can this year place-take  
 ‘The book fair can take place this year.’
- b. *Die Buchmesse finde-t dieses Jahr statt.*  
 the book.fair take-s this year place  
 ‘The book fair takes place this year.’

Since the definition is not meant to be a fuzzy concept or prototype (see (iii) in §1 above), this also means that constructions with linking elements as in (10) are not included.

- (10) not compounds:
- |             |                      |                  |                          |
|-------------|----------------------|------------------|--------------------------|
| a. German   | <i>Liebe-s-brief</i> | [love-LK-letter] | ‘love letter’            |
| b. English  | <i>bird’s nest</i>   |                  |                          |
| c. French   | <i>chemin de fer</i> | [way of iron]    | ‘railway’                |
| d. M. Greek | <i>vrox-ó-nero</i>   | [rain-LK-water]  | ‘rainwater’ <sup>7</sup> |

<sup>6</sup> Since such compound-like constructions are not compounds, I do not know of any way of defining them cross-linguistically. Thus, I cannot (at the moment) make any general statements about the forms that have been called “phrasal compounds”. Thus, “they” does not have a clear reference in the last sentence of this section.

<sup>7</sup> This example is from Ralli (2013: 47); the relevant words are *vroxí* ‘rain’ and *neró* ‘water’.

In Indo-European languages, such compound-like constructions are not uncommon, and at least in German, they have always been included in the (language-particular) category of Noun-noun Compounds. In a classic article, Benveniste (1966) said that forms of the type *chemin de fer* are the true compounds of French. However, the subsequent literature has not adopted this way of talking about such Romance expressions. They have recently been called “phrasal lexemes” (Masini 2009) or *binominal lexemes* (Masini et al. 2023; Pepper 2023), defined in terms of the classifying or naming function of such forms. However, the term *compound* is generally defined in a strictly formal way (see, e.g., the definitions listed by Scalise & Vogel 2010: 5), and this tradition is followed here. As a result, not only Romance expressions containing a preposition of the *de/di* type are excluded, but also Germanic expressions that include a genitive-type marker or other additional markers that are not part of either of the roots.

Another example of a compound-like construction that is not a compound is shown in (11). One might consider treating *lí-hūn* [leave-marriage] ‘divorce’ as an incorporation (see Wang 2022), but some adverbials can intervene between the two parts of this composite expression. It is thus more similar to English expressions like *take part*, which cannot be considered incorporations either because some affixes can intervene (*take-s part, tak-ing part*).

(11) Mandarin Chinese

- a. *lí-hūn* (离婚) [leave-marriage] ‘divorce’  
 b. *lí-le-liǎng-cì-hūn* (离了两次婚) [leave-PRF-two-time-marriage] ‘divorced twice’

## 6. Roots in compounds cannot be expanded

We now get to the key property of compound constructions that distinguishes them from “syntactic” constructions: nonexpandability. A widely shared intuition is that compounds are formed “morphologically”, and they are typically included under the heading of “word formation”. But how can this be translated into a straightforward criterion that can be applied uniformly to all languages? This question is often formulated as the question of distinguishing between compounds and “phrases” (e.g. Schlücker & Plag 2011; Ralli 2013: 243-268; Cetnarowska 2019: 15-44; Gebhardt 2023: 136-140).

But how do we recognize a “phrase”? In sentences such as *cats like milk*, or *he lacks courage*, the subject and object forms are typically treated as nominal phrases rather than simply words, but this cannot be seen in these particular sentences. The reason why we say that *cats, milk, he* and *courage* occupy phrasal slots in a construction of the type [NP V NP] is that they can be expanded by articles and adjectival or nominal modifiers, as in *small cats like my neighbour’s milk*, or *he lacks the necessary courage*. For adjective-noun compounds, the impossibility of modification by degree adverbs has often been noted in the literature; for example, Finkbeiner & Schlücker (2019: 10) observe that the first element in German *Alt-bau* [old-building] cannot be expanded by a degree adverb (*\*sehr Alt-bau* ‘very old building’).

For Modern Greek, Ralli (2013: 21) notes that the phrase *áyria yáta* ‘wild cat’ can be expanded by a noun phrase modifier or by coordination, while the constituents of the compound *ayrió-yata* ‘wildcat’ cannot be expanded in this way (though it is of course possible to modify the entire compound, as in *meyáli ayrió-yata* ‘big wildcat’).<sup>8</sup>

<sup>8</sup> It should be noted that the expansion criterion does not distinguish fixed phrases from fixed compounds, but only freely formed phrases from freely formed compounds. For example, Schlücker & Plag (2011: 1541) observe that the German fixed phrase *schwarzes Brett* [black board] ‘bulletin board’ cannot be

- (11)a. *áyria yáta*  
wild cat  
'wild cat'
- b. *i áyria tis Mariás i yáta*  
the wild of Maria the cat  
'Maria's wild cat'
- c. *áyria ke meýali yáta*  
wild and big cat  
'wild and big cat'
- (12)a. *ayrió-yata*  
wild-cat  
'wildcat (*Felis silvestris*)'
- b. *\*ayrio-mavrió-yata*  
wild-black-cat  
'wild black cat'
- c. *\*poli-ayrió-yata*  
very-wild-cat  
'very wild cat'

The definition in (1) does not mention articles or coordination because these do not exist in all languages, and the definition can make use only of universally applicable concepts. As adjectival modifiers and possessive modifiers of nouns occur in all languages, and likewise degree modifiers of adjectives (presumably) occur in all languages, they are suitable for the general definition in (1). For verbs, many languages have adjectival modifiers (e.g. English *walk slowly*, German *schön singen* 'sing beautifully'), and maybe one should add other types of expansions to the definition in (1) to distinguish verb roots in compounds from verbs occurring in "phrases".

Thus, the criterion of expandability allows us to distinguish between compounds and what has traditionally been called "phrases" without requiring a definition of "phrase". We will see below that this criterion also allows us to distinguish incorporation from constructions that are not incorporation.

## 7. Compounds need not have a naming or generic function

The most typical nominal compounds denote established concepts that are known to the language users independently of the compound construction. For example, *summer vacation* in (13a) is an established concept in modern Western culture, while (13b) refers to a specific instance of a summer vacation experienced by a specific person. Schlücker & Plag (2011) say that that compounds are "inherently suitable for kind reference (or "naming"), due to their status as word formation entities", and this seems to be a widespread view.

---

expanded to *\*sehr schwarzes Brett* [very black board], but this is for semantic reasons. (Thus, Gebhardt's *\*very bluebird* is not really a convincing argument, because *bluebird* is a fixed idiomatic compound.)

- (13) a. *summer vacation* (naming compound)  
 b. *last summer's vacation in Czechia* (specific phrase)

However, compounds need not have a naming function or refer to kinds rather than specific referents. Many languages allow ad hoc compounds such as *office vacation* in (14a), which is not an established concept but could refer to an unusual kind of vacation (e.g. a vacation during which an agricultural worker gets to use an empty office and can relax by working on their memoir for a few weeks). And in addition, compounds need not be generic (kind-referring), but the modifying root can refer to a specific person, as in *Miterrand interview* in (14b).

- (14) a. *office vacation* (ad hoc compound)  
 b. *the Mitterrand interview* (specific modifying root)

As I noted above, *compound* and *incorporation* are best defined in strictly formal terms, and frequently occurring functions of compound or incorporation constructions are best characterized by different terms. This is also the conclusion of Schlücker & Plag (2011), who note that not only adjective-noun compounds (such as *Rotwein* ‘red wine’) can have a naming function in German, but also adjective-noun phrases (such as *großer Zeh* ‘big toe’ or *bunter Abend* [colorful evening] ‘evening of music and entertainment’). Croft (2022: 141-142) calls the naming or generic function *typefying*, and he notes that many languages do not use compounds for typefying constructions (e.g. preferring binominals such as French *chemin de fer* ‘railway’, cf. §5 above). A compound construction is thus a construction-strategy, not a construction-function.

## 8. Some further properties of the definition

Before moving on to incorporation constructions, I will briefly comment on five further topics: (i) the specification of two roots; (ii) the treatment of “synthetic compounds”; (iii) the treatment of “neoclassical compounds”; (iv) the possible presence of an external compound marker; and (v) the order of the two roots.

First, one may ask why the definition restricts compound constructions to binary combinations of roots. There is no strong reason for this, but it makes the definition more concrete than saying “several roots”, and the possible ternary (or  $n$ -ary for  $n > 3$ ) combinations that are excluded in this way seem to be marginal. One might say that a compound such as *Brazil-Argentina game* is ternary (because *Brazil-Argentina* is not an element that undisputably exists outside of the compound), but such compounds are uncommon, and semantically they are not ternary.

Second, what about “synthetic compounds” such as German *hart-herz-ig* [hard-heart-y] ‘uncharitable, hard-hearted’ (based on *hart* and *Herz*), or Spanish *medi-ev-al* [middle-age-al] ‘medieval’ (based on *medi(-o)* and *ev(-o)*)? These include two roots, but crucially an additional affix, so they are not compounds according to the present definition.

Third, “neoclassical compounds” of European languages such as *geo-logy*, *bio-graphy*, *demo-cracy* do not count as compounds because their components are not roots (in that they do not occur in free forms without another root-like form; see (7)).

Fourth, as noted by Trips & Kornfilt (2015), Turkish compound-like forms include a marker at the end of the second root, e.g. *müzik festival-i* ‘music festival’. Such markers go beyond what is specified in the definition, so they do not count as compounds. One might ask what they are if they are not compounds, because they are not ordinary phrases



either. But this is not a question that I address here. Not every language-particular construction must fall under some well-known grammatical concept, and there are many *sui generis* phenomena in languages.

And fifth, the definition does not specify the order in which the two roots occur, so theoretically, the order could be free. I do not know about compound constructions with free order, but if there are no such constructions, this may be regarded as a testable universal, and it need not be specified in the definition.

## 9. More on incorporation

In this paper, incorporation is defined as a subtype of compounding. The definition in (2) above is repeated here for convenience.

### (2) incorporation construction

An incorporation construction is an event-denoting verb-noun compound construction in which the noun occupies an argument slot of the verb and occurs in a position where nominal patient arguments cannot occur.

Analogously to the distinction between *compound construction* and *compound* (in §3 above) we can distinguish between an *incorporation construction* and an *incorporation* (as a kind of form that instantiates an incorporation construction), though this is not usual in the earlier literature.<sup>9</sup>

That the verb-noun compounds that are called incorporations refer to the event denoted by the verb is the first important condition, needed to exclude compounds with verb modifiers such as English *push-cart* or German *Wasch-maschine* ‘washing machine’, as well as exocentric compounds such as French *tire-bouchon* [pull-cork] ‘corkscrew’. In addition, the noun must occupy an argument position of the verb (mostly that of patient argument).

Let us briefly look at a few definitions of (*noun*) *incorporation* in the literature, to see some of the specificities of the present definition:<sup>10</sup>

- (16) a. “Noun incorporation is the compounding of a noun stem and a verb (or adjective) to yield a complex form that serves as the predicate of a clause.” (Gerds 1998)
- b. [noun incorporation:] “a construction in which a noun stem is combined with a verb to form a new, morphologically complex verb” (Mithun 2000: 916)
- c. “we will define noun incorporation as instances where a bare or reduced nominal displays a close linear linguistic relation with a verb, through either morphology or strict adjacency, and where the head of this unit is clearly

<sup>9</sup> Note that *incorporation* is also often used to denote an abstract process, analogously to *compounding* (= the process of forming a compound). In the present paper, I avoid such process nouns.

<sup>10</sup> There is no need for the modifier “noun” because incorporation (as defined here) is by definition noun incorporation. Some authors have talked about “adverb incorporation” or incorporation of other elements, but these phenomena are clearly marginal and can safely be excluded. Gerds (1998) basically defines “incorporation” as coextensive with compound formation and thus treats noun incorporation as a special case (see also Haugen 2015: 414). (The case of “incorporated pronouns” is a different matter, because it is clearly completely unrelated: When person indexes are affixed, they are sometimes said to be “incorporated”, in a usage that goes back to Sapir (1911: 250) or earlier, but this terminological usage is confusing.)

verbal in its distribution or marking.” (Johns 2017: §1)

- d. “as a starting point, we can say that it refers to a grammatical phenomenon whereby a nominal element, usually with an internal thematic role, forms some kind of unit with a verbal element, and together they serve as the verb or predicate of a sentence.” (Massam 2017: §1)
- e. “Incorporation can be described as the inclusion of one lexical element in another lexical element such that they together constitute a single word.” (Olthof 2020: 131)
- f. “Noun incorporation prototypically consists of a verbal compound consisting of a verbal root and a nominal root.” (Barrie & Mathieu 2020: 265)

Most of these definitions are largely compatible with the present definition, but some of them make use of concepts that are not needed here: “word” in (16e), “morphology” in (16b, c), “predicate” in (16a, d), “head” in (16c), “lexical element” in (16e). I do not know how to define these terms, so the present definition of *incorporation* instead makes reference to well-defined notions such as *compound* (instead of “morphology”), *root* (instead of “lexical element”), and “event-denoting” (instead of “head”).

Incorporation became well-known in the 1980s because of a prominent controversy between Mithun (1984; 1986) representing a “lexical” view of incorporation and Sadock (1986) representing a “syntactic” view, and because of Baker’s (1988) use of the term “incorporation” for an abstract syntactic operation (see Haugen 2015 for an account of key aspects of the debates). Like many hotly debated issues in linguistics, it has not been resolved, and if one does not presuppose a distinction between “syntax” and “lexicon”, the controversy loses much of its interest.

One issue has been the question whether an incorporated noun can be referential and serve as an antecedent for a subsequent anaphoric construction. Sadock (1986) notes that this is possible in Greenlandic, as seen in (17a-b).

(17) Greenlandic

- a. *Luutiviup assut qusanartumik qaanniorpaa,*  
 Luutivik-p      assut      qusanartoq-mik      qajaq-lior-paa  
 Luutivik-ERG    very      beautiful-INS      kayak-make.for-INDIC.3SG>3SG  
 ‘Luutivik made him a very beautiful kayak,’
- b. *unnermillu amertillugu.*  
 unneq-mik=lu                      amertit-lugu  
 white.sealskin-INS=CONJ      cover-CONTEMP.3SG  
 ‘and covered it with white sealskin.’ (Sadock 1986: 23)

Sadock argued that this shows that at least some languages have a kind of noun incorporation that is syntactic, and that there not all incorporations are “lexical” (as claimed by Mithun 1984). However, this observation does not affect the current discussion, because (i) there is no claim that parts of compounds cannot be specific or referential (see §7), and (ii) the Greenlandic construction discussed by Sadock does not count as incorporation by the criteria of the present definition. This is because Greenlandic *-lior/-nior* ‘make for’ (as in *qaan-nior-paa* ‘made a kayak for him’) is a suffix rather than a root. As specified in the definition in (5), a root “can occur as part of a free form without another contentful morph”, but this is not the case for *-lior/-nior*,

which always occurs following a noun root and is thus a derivational suffix, not a root. Eskimo languages are unusual in having a large number of derivational suffixes (often called “postbases”), and it is also unusual that the noun roots in such derived verbs can be referential, but these are not cases of incorporation.

Likewise, derivational suffixes that mean ‘make’ or ‘have’ in Uto-Aztecan languages are not roots, although they have been treated in the context of incorporation by Hale & Keyser (2002) and Haugen (2008) (see also Haugen 2015: §4). Two examples from O’odham are given in (18).

- (18) Tohono O’odham (Uto-Aztecan)
- |    |             |          |               |                 |
|----|-------------|----------|---------------|-----------------|
| a. | <i>hoa</i>  | ‘basket’ | <i>hoa-t</i>  | ‘make a basket’ |
| b. | <i>si:l</i> | ‘saddle’ | <i>si:l-t</i> | ‘make a saddle’ |

Such denominal suffixes are also discussed by Gerds & Marlett (2008) and contrasted with incorporation. Thus, they note that the Halkomelem element *tx<sup>w</sup>*- ‘buy’ cannot occur on its own, but only in combination with a noun root as in (19a). Thus this construction is not an incorporation construction.

- (19) Halkomelem (Salishan; Gerds & Marlett 2008)

- |    |                        |            |                               |                      |                |
|----|------------------------|------------|-------------------------------|----------------------|----------------|
| a. | <i>Ni?</i>             | <i>cən</i> | <i>tx<sup>w</sup>-səplil.</i> |                      |                |
|    | AUX                    | 1.SBJ      | VBL-bread                     |                      |                |
|    | ‘I bought bread.’      |            |                               |                      |                |
| b. | <i>*Ni?</i>            | <i>cən</i> | <i>tx<sup>w</sup>-ət</i>      | <i>k<sup>w</sup></i> | <i>səplil.</i> |
|    | AUX                    | 1.SBJ      | buy-TR                        | DET                  | bread          |
|    | ‘I bought some bread.’ |            |                               |                      |                |
| c. | <i>Ni?</i>             | <i>cən</i> | <i>ʔiləq-ət</i>               | <i>k<sup>w</sup></i> | <i>səplil.</i> |
|    | AUX                    | 1.SBJ      | buy-TR                        | DET                  | bread          |
|    | ‘I bought some bread.’ |            |                               |                      |                |

Finally, the definition in (2) specifies that the incorporated noun must occur in a position different from the position of the argument when it is a full nominal. For example, in Turkish, bare caseless nouns in object function can only occur directly in front of the verb, as in (20a), in contrast to accusative-marked objects, as seen in (20b).

- (20) Turkish (Aydemir 2004: 465-466)

- |    |  |                                      |
|----|--|--------------------------------------|
| a. | <i>Yasemin anahtar kaybet-ti.</i>                            | <i>(*Anahtar Yasemin kaybet-ti.)</i> |
|    | Yasemin key lose-PST   |                                      |
|    | ‘Yasemin lost keys (or a key).’                              |                                      |
| b. | <i>Anahtar-ı Yasemin kaybet-ti.</i>                          |                                      |
|    | key-ACC Yasemin lose-PST                                     |                                      |
|    | ‘Yasemin lost the key.’ (OR: ‘The key was lost by Yasemin.’) |                                      |

So does (20a) represent a case of noun incorporation, as hinted by Aydemir (2004)? I would say that it does not, because the definite nominal can occur in the same position (*Yasemin anahtarı kaybet-ti*, which is basically equivalent to (20b)). Of course, in an abstract analysis, linguists may set up different structural positions for different types of objects (as they often do), but comparative concepts for typology need to rely on criteria that can be read off the concrete forms.

## 10. Shared-core definitions of concepts for general linguistics

For some or many readers, one of the most surprising or perhaps irritating aspects of the current proposal is that several types of constructions or expressions that are typically treated as compounds are not covered by the definitions. In particular, the exclusion of compound-like forms with composite components and of “phrasal compounds” (§4), as well as the exclusion of “synthetic compounds” (§8), may seem unintuitive. However, it must be kept in mind that such “surprising” properties are a necessary feature of all definitions of general (comparative) grammatical terms, because they are defined in a way different from language-particular categories. General concepts cover the SHARED CORE of the extension of the comparable categories of different.

For example, the English Present Tense and the German Present Tense are not coextensive, because only the German Present Tense can regularly be used for future time reference as well (e.g. *wir kommen morgen* ‘we will come tomorrow’), but they both match the comparative concept “present tense”. Or the Turkish Dative case and the Russian Dative case are not coextensive, because only the Turkish Dative case can be used for spatial goals as well (e.g. *Moskova’ya* ‘to Moscow’), but they both match the comparative concept “dative case”. Or the French Feminine gender class is not coextensive with the English Feminine gender class because only the French Feminine gender class includes many inanimate nouns (e.g. *la lune* ‘the moon’), but they both match the comparative concept “feminine gender class”.

Similarly, German Compounds and Chinese Compounds overlap in a core set of phenomena that match the definition in (1), but for both languages, there are language-particular criteria that lead researchers to include more phenomena. This is not different from the language-particular criterion for French Feminine gender, which is defined by the article *la*, not by the meanings of the nouns.

## 11. Conclusion

To conclude, the definitions of *compound* and *incorporation* proposed in this paper make it possible to compare these phenomena across languages without presupposing a distinction between syntax and morphology, or definitions of “word” or “phrase”. A key concept here is that of expansion by full nominal, adjectival or degree modifiers, which yields results that earlier linguists often tried to describe in terms of a “word–phrase” distinction or a “syntax–morphology” distinction.

In what sense does this constitute progress? There is no empirical contribution in this paper, but the definitions suggested here show that there is no reason to resign oneself to a situation where we do not have a definition shared by the discipline. Quite a few authors have expressed rather sobering or negative views about definitions, e.g.

- (20) a. “There is no overall agreement on such basic issues as the definition of a compound. Accordingly, there can be no agreement on whether compounding is a linguistic universal or not.” (Bauer 2017: 1-2)
- b. “It’s difficult to classify compounds and phrases into two distinct morphosyntactic structures.” (Gebhardt 2023: 140)
- c. “Noun incorporation can be defined in a number of ways.” (Johns 2017: §1)

- d. “There is a lot of disagreement about exactly what constitutes noun incorporation.” (Massam 2017: §1)

It seems that the main reason why linguists often disagree about the precise meanings of even rather basic general terms is that they think of these terms as denoting aspects of nature, rather than concepts created by linguists in order to compare languages. Linguists often seem to assume that a good definition of a term is something that will result from a definitive theory of the domain in question, or at least that our definitions will get better as our theories get better.

However, if general terms of linguistics denote comparative concepts rather than natural kinds, there is no reason to wait for better theories. There does not seem to be any tendency for linguists’ general-theoretical and methodological views and preferences to converge over the decades, so it is worth exploring the possibility of finding definitions that could work for everyone because they do not rely on controversial theoretical views. It is hoped that the present definitions have this property, so that even if readers are not happy to adopt them, they will appreciate the fact that the definitions do not rely on other basic concepts (such as “syntax vs. morphology”, or “head”, or “phrase”) that are themselves controversial or poorly understood.

## References

- Aydemir, Yasemin. 2004. Are Turkish preverbal bare nouns syntactic arguments? *Linguistic Inquiry* 35(3). 465–474.
- Baker, Mark C. 1988. *Incorporation: A theory of grammatical function changing*. Chicago: University of Chicago Press.
- Baker, Mark C. & Aranovich, Roberto & Golluscio, Lucía A. 2005. Two types of syntactic noun incorporation: Noun incorporation in Mapudungun and its typological implications. *Language*. Linguistic Society of America 81(1). 138–176.
- Barrie, Michael & Mathieu, Eric. 2020. Noun incorporation and polysynthesis. In Siddiqi, Daniel & Barrie, Michael & Gillon, Carrie & Haugen, Jason D. & Mathieu, Eric (eds.), *The Routledge handbook of North American languages* (Routledge Handbooks in Linguistics), 253–274. New York, NY: Routledge.
- Bauer, Laurie. 2017. *Compounds and compounding* (Cambridge Studies in Linguistics). Cambridge: Cambridge University Press.
- Benveniste, Émile. 1966. Formes nouvelles de la composition nominale. *Bulletin de la Société de Linguistique de Paris* 61(1). 82–95.
- Cetnarowska, Bożena. 2019. *Compound nouns and phrasal nouns in English and Polish* (Prace Naukowe Uniwersytetu Śląskiego w Katowicach). Katowice: Wydawnictwo Uniwersytetu Śląskiego.
- Croft, William. 2022. *Morphosyntax: Constructions of the world’s languages*. Cambridge: Cambridge University Press.
- Downing, Pamela. 1977. On the creation and use of English compound nouns. *Language*. Linguistic Society of America 53(4). 810–842. (doi:10.2307/412913)
- Dressler, Wolfgang U. 2006. Compound types. In Libben, Gary & Jarema, Gonia (eds.), *The representation and processing of compound words*, 23–44. Oxford: Oxford University Press.
- Evans, Nicholas. 1996. The syntax and semantics of body part incorporation in Mayali. In Chappell, Hilary & McGregor, William (eds.), *The grammar of inalienability*, 65–110. Berlin: Mouton de Gruyter.
- Finkbeiner, Rita & Schlücker, Barbara. 2019. Compounds and multi-word expressions in the languages of Europe. In Schlücker, Barbara (ed.), *Complex lexical units: Compounds and multi-word expressions*, 1–44. Berlin: De Gruyter. (doi:10.1515/9783110632446-001)

- Gebhardt, Lewis. 2023. *The study of words: An introduction*. London: Routledge. (doi:10.4324/9781003030188)
- Gerdts, Donna B. 1998. Incorporation. In Spencer, Andrew & Zwicky, Arnold M. (eds.), *The handbook of morphology*. Oxford: Blackwell.
- Gerdts, Donna B. & Marlett, Stephen A. 2008. Introduction: The form and function of denominal verb constructions. *International Journal of American Linguistics* 74(4). 409–422. (doi:10.1086/595571)
- Hale, Ken & Keyser, Samuel Jay. 2002. *Prolegomenon to a theory of argument structure*. Cambridge, MA: MIT Press.
- Haspelmath, Martin. 2011. The indeterminacy of word segmentation and the nature of morphology and syntax. *Folia Linguistica* 45(1). 31–80. (doi:10.1515/flin-2017-1005)
- Haspelmath, Martin. 2016. The serial verb construction: Comparative concept and cross-linguistic generalizations. *Language and Linguistics* 17(3). 291–319. (doi:http://doi.org/10.1177/2397002215626895)
- Haspelmath, Martin. 2018. How comparative concepts and descriptive linguistic categories are different. In Van Olmen, Daniël & Mortelmans, Tanja & Brisard, Frank (eds.), *Aspects of linguistic variation: Studies in honor of Johan van der Auwera*, 83–113. Berlin: De Gruyter Mouton. (https://zenodo.org/record/3519206)
- Haspelmath, Martin. 2021. General linguistics must be based on universals (or nonconventional aspects of language). *Theoretical Linguistics* 47(1–2). 1–31. (doi:10.1515/tl-2021-2002)
- Haspelmath, Martin. 2023. On what a construction is. *Constructions (to appear)*. (https://ling.auf.net/lingbuzz/007200)
- Haspelmath, Martin. 2024. Inflection and derivation as traditional comparative concepts. *Linguistics (to appear)*.
- Haugen, Jason D. 2008. *Morphology at the interfaces: Reduplication and noun incorporation in Uto-Aztecan*. Amsterdam: Benjamins.
- Haugen, Jason D. 2015. Incorporation. In Müller, Peter O. & Ohnheiser, Ingeborg & Olsen, Susan & Rainer, Franz (eds.), *Word-Formation: An International Handbook of the Languages of Europe, Volume 1*, 413–434. Berlin: De Gruyter Mouton. (doi:10.1515/9783110246278)
- Johns, Alana. 2017. Noun incorporation. In Everaert, Martin & van Riemsdijk, Henk (eds.), *The Wiley Blackwell companion to syntax*. Wiley Blackwell. (https://doi.org/10.1002/9781118358733.wbsyncom093)
- Lovstrand, Joseph. 2021. Serial verb constructions. *Annual Review of Linguistics* 7(1). 109–130. (doi:10.1146/annurev-linguistics-031920-115317)
- Marchand, Hans. 1969. *The categories and types of present-day English word-formation: A synchronic-diachronic approach*. Munich: C.H. Beck.
- Masini, Francesca. 2009. Phrasal lexemes, compounds and phrases: A constructionist perspective. *Word Structure*. Edinburgh University Press 2(2). 254–271. (doi:10.3366/E1750124509000440)
- Masini, Francesca & Mattioli, Simone & Pepper, Steve. 2023. Exploring complex lexemes cross-linguistically. In Pepper, Steve & Masini, Francesca & Mattioli, Simone (eds.), *Binominal lexemes in cross-linguistic perspective: Towards a typology of complex lexemes*, 1–19. Berlin: De Gruyter Mouton.
- Massam, Diane. 2017. Incorporation and pseudo-incorporation in syntax. *Oxford Research Encyclopedia of Linguistics*. (doi:10.1093/acrefore/9780199384655.013.190)
- Mithun, Marianne. 1984. The evolution of noun incorporation. *Language* 60(4). 847–894.
- Mithun, Marianne. 1986. On the nature of noun incorporation. *Language* 62(1). 32–37.
- Mithun, Marianne. 2000. Incorporation. In Booij, Geert & Lehmann, Christian & Mugdan, Joachim (eds.), *Morphology: A handbook on inflection and word formation, vol. 1* (Handbücher Zur Sprach- Und Kommunikationswissenschaft), 916–929. Berlin: Mouton de Gruyter.
- Mithun, Marianne. 2010. Constraints on compounds and incorporation. In Scalise, Sergio & Vogel, Irene (eds.), *Cross-disciplinary issues in compounding*, 37–56. Amsterdam: Benjamins.

- Olsen, Susan. 2015. Composition. In Müller, Peter O. & Ohnheiser, Ingeborg & Olsen, Susan & Rainer, Franz (eds.), *Word-formation: An international handbook of the languages of Europe*, vol. Vol. 1. Berlin: De Gruyter Mouton. (doi:10.1515/9783110246254-017)
- Olthof, Marieke. 2020. Formal variation in incorporation: A typological study and a unified approach. *Linguistics*. De Gruyter Mouton 58(1). 131–205. (doi:10.1515/ling-2019-0036)
- Pepper, Steve. 2023. Defining and typologizing binominal lexemes. In Pepper, Steve & Masini, Francesca & Mattioli, Simone (eds.), *Binominal lexemes in cross-linguistic perspective: Towards a typology of complex lexemes*, 23–72. Berlin: De Gruyter Mouton.
- Ralli, Angela. 2013. *Compounding in Modern Greek*. New York: Springer.
- Sadock, Jerrold M. 1986. Some notes on noun incorporation. *Language* 62(1). 19–31.
- Sapir, Edward. 1911. The problem of noun incorporation in American languages. *American Anthropologist* 13(2). 250–282.
- Scalise, Sergio & Vogel, Irene (eds.). 2010. *Cross-disciplinary issues in compounding*. Amsterdam: Benjamins.
- Schlücker, Barbara. 2023. Compounding and linking elements in Germanic. *Oxford Research Encyclopedia of Linguistics* 2023. (doi:10.1093/acrefore/9780199384655.013.954)
- Schlücker, Barbara & Plag, Ingo. 2011. Compound or phrase? Analogy in naming. *Lingua* 121(9). 1539–1551. (doi:10.1016/j.lingua.2011.04.005)
- Spencer, Andrew. 1995. Incorporation in Chukchi. *Language* 71(3). 439–489.
- Trips, Carola & Kornfilt, Jaklin. 2015. Typological aspects of phrasal compounds in English, German, Turkish and Turkic. *STUF - Language Typology and Universals* 68(3). 281–321. (doi:10.1515/stuf-2015-0015)
- Velázquez-Castillo, Maura. 1996. *The grammar of possession: Inalienability, incorporation and possessor ascension in Guarani* (Studies in Language Companion Series). Amsterdam: Benjamins.
- Wälchli, Bernhard. 2015. Co-compounds. In Müller, Peter O. & Ohnheiser, Ingeborg & Olsen, Susan & Rainer, Franz (eds.), *Word-Formation: An International Handbook of the Languages of Europe, Volume 1*, 707–727. Berlin: De Gruyter Mouton. (doi:10.1515/9783110246278)
- Wang, Yong. 2022. From syntax to morphology: Noun-incorporation in Chinese. *Studies in Language* 46(4). 872–900. (doi:10.1075/sl.21015.wan)
- Wurzel, Wolfgang Ullrich. 1996. Morphologischer Strukturwandel: Typologische Entwicklungen im Deutschen. In Lang, Ewald & Zifonun, Gisela (eds.), *Deutsch - typologisch*, 492–524. Berlin: Mouton de Gruyter. (doi:10.1515/9783110622522-022)