Compound words, incorporation, and serialization

Martin Haspelmath
Max Planck Institute for Evolutionary Anthropology

1. General linguistics: definitions and universals

– particular linguistics: analyzing language structures

– general linguistics: finding (and explaining) universals identified
on the basis of universally applicable definitions

p-categories: defined in language-particular ways
g-concepts: defined in a uniform way

G-concepts (= comparative concepts) are the basis for language **comparison** and language **universals**, and these are the basis for **general linguistics** (Haspelmath 2021).

Linguists often think that one needs a definitive theory before one can have good definitions of terms, so we need to be resigned to not having them:

(1)   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.   There is no clear or general definition of compound" (Schlücker 2023: §2.2)

   d.   "Noun incorporation can be defined in a number of ways." (Johns 2017: §1)

   e.   "There is a lot of disagreement about exactly what constitutes noun incorporation." (Massam 2017: §1)

Here I propose a different perspective: that our general technical terms are part of our **methodology**, not (necessarily) part of our theoretical understanding – hence we do not need a definitive theory (merely a set of **uncontroversial basic terms**).

2. Defining *compound construction* and *incorporation construction*

(1) Definition 1: **compound construction**
A compound construction is a construction consisting of two slots for roots that occur adjacent and that cannot be expanded by full nominal, adjectival or degree modifiers.

(2) Definition 2: **incorporation construction**
An incorporation construction is a verb-noun compound construction which refers to the event denoted by the verb and in which the noun occupies an argument slot of the verb and occurs in a position where nominal patient arguments cannot occur.

(3) some compound forms
a. German	*Auto-bahn*	[car-way]	'expressway'
b. French	*tire-bouchon*	[pull-cork]	'corkscrew'
c. Chinese	飛機 fēi-jī	[fly-machine]	'airplane'
d. Mwotlap	*tit teñteñ*	[punch cry]	'make (someone) cry by punching'

(4) some noun-incorporating verb forms
a. Mohawk
  *wak-**tsitsia**-ientho-on*
  1SG-flower-plant-ST
  'I planted flowers' (Mithun 2010)

b. Bininj Gunwok
  *ba-**warde**-jobge-ng*
  3PL-rock-split-PST
  'they split the rock' (Evans 1996: 65)

c. Guarani
  *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)

Notable properties of these definitions:

(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.

German Compounds are defined with respect to a special stress pattern:

compound *Rótwein* 'red wine', vs. *ròter Wéin* 'wine which is red'
(cf. *Langewéile*: not a compound)

But this criterion plays no role in many other languages.

3. Compounds consist of roots (not stems or words)

"compounds consist of two words" (Marchand 1969) or two stems (Schlücker 2023)

Olsen (2015: 364): compounding involves the combination of "two or more lexemes (roots, stems, or freely occurring words)".

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

(5) Definition 3: **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.

Thus:

– 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,

– combinations involving pronouns (e.g. English *him-self*) or adpositions (e.g. *on-to*) are not regarded as compounds

Wrong: two **"words"** or **"lexemes"** because words often include inflectional affixes:

    cf. Latin adjective-noun compound *magn-anim-(us)* 'great-spirited'
            root *magn-* 'great' (nominative singular *magn-us*, accusative singular *magn-um*, dative plural *magn-is*, and so on).

Or maybe: "compounds consist of stems (rather than roots)?" cf.

(6)   English   *free-dom day*
       German   *Gas-heiz-ung* [gas-heat-ing]   'gas heating'
       Hungarian   *hó-es-és*   [snow-fall-NMLZ]   'snowfall'

But what exactly is a "stem"? a combination of a root and derivational affixes?

(There is no generally accepted definition of "derivational" vs. "inflectional", but see Haspelmath 2024 for a recent proposal).

Fiinally: **phrases** cannot be compound members – so-called "phrasal compounds" of the type *chicken and egg situation* do not fall under the current definition (cf. Trips & Kornfilt 2015)

4. The roots are adjacent

Auto-fahren is not a compound:

(7)   a. *Wir möchten   heute   Auto   fahren.*
        we   would.like   today   car   drive
        'We would like to go by car today.'

b. *Wir fahren heute Auto.*
   we   drive   today   car
   'We go by car today.'

Constructions with linking elements as in (8) are not included.

(8) a. German *Liebe-s-brief* [love-LK-letter] 'love letter'
    b. English *bird's nest*
    c. French *chemin de fer* [way of iron] 'railway'
    d. M. Greek *vrox-ó-nero* [rain-LK-water] 'rainwater' (Ralli 2013)

Benveniste (1966): forms of the type *chemin de fer* are the true compounds of French.

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.

Cf. also:

(9) 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'

– similar to English expressions like *take part*, which cannot be considered incorporations either (*take-s part, tak-ing part*).

5. Roots in compounds cannot be expanded

A widespread view: "compounds are not formed syntactically, but morphologically, as part of word formation".

But how do we distinsguish between syntax and morphology, or between **compounds** and **"phrases"** (e.g. Schlücker & Plag 2011; Ralli 2013: 243-268; Cetnarowska 2019: 15-44; Gebhardt 2013: 136-140), or between?

Consider examples like

  cats like milk
  he lacks courage: [NP V NP]

The reason why we say that *cats*, *milk*, *he* and *courage* **occupy phrasal slots** here is that they can be expanded by articles and adjectival or nominal modifiers, as in

  small cats like my neighbour's milk
  he lacks the necessary courage

In adjective-noun compounds, the adjective cannot be modified by degree adverbs, e.g.

Gebhardt (2023: 139): English *bluebird* cannot be expanded to **very bluebird*)

Modern Greek (Ralli 2013: 21)

(10) a. *áɣria ɣáta*
     wild cat
     'wild cat'

   b. *i   áɣria tis Marías i   ɣáta*
      the wild   of Maria   the cat
      'Maria's wild cat'

   c. *áɣria ke   meɣáli ɣáta*
      wild and big   cat
      'wild and big cat'

(11) a. *aɣrió-ɣata*
     wild-cat
     'wildcat (*Felis silvestris*)'

   b. **aɣrio-mavrió-ɣata*
      wild-black-cat
      'wild black cat'

   c. **poli-aɣrió-ɣata*
      very-wild-cat
      'very wild cat'

The criterion of **expandability** allows us to distinguish between compounds and what has traditionally been called "phrases" without requiring a definition of "phrase".

6. Compounds need not have a naming or generic function

Schlücker & Plag (2011): compounds are "inherently suitable for kind reference (or "naming"), due to their status as word formation entities" (this seems to be a widespread view)

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

But: compounds need not have a naming function or refer to kinds rather than specific referents!

– Many languages allow ad hoc compounds

– Compounds need not be generic (kind-referring), but the modifying root can refer to a specific person:

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

– *compound* and *incorporation* are best defined in **strictly formal terms**

– frequently occurring functions of compound or incorporation constructions are best characterized by different terms (e.g. binominal, Pepper 2023)

– this is in line with Schlücker & Plag (2011), who note that not only **adjective-noun compounds** (such as *Rotwein* 'red wine') can have a naming function, but also **adjective-noun phrases** (such as *großer Zeh* 'big toe' or *bunter Abend* [colorful evening] 'evening of music and entertainment').

7. More on incorporation

– incorporation is defined as a subtype of compounding here:

(2) Definition 2: **incorporation construction**
    An incorporation construction is a verb-noun compound construction which refers to the event denoted by the verb and in which the noun occupies an argument slot of the verb and occurs in a position where nominal patient arguments cannot occur.

– incorporations refer to **the event denoted by the verb** – this excludes English *push-cart* or German *Wasch-maschine* 'washing machine', as well as exocentric compounds such as French *tire-bouchon* [pull-cork] 'corkscrew'

Some earlier definitions:

(14) 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." (Gerdts 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 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)

Concepts not needed here: "predicate" in (14a, d), "morphology" in (14b, c), "head" in (14c), "lexical element" in (14e).

Here: *compound*, *root*, "event denoted by the verb

Can incorporations be syntactic? Sadock (1986) vs. Mithun (1984)

(15) 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)

Denominal verb affixes do not involve incorporation (Gerdts & Marlett 2008; cf. Baker 1988):

(16) Halkomelem (Salishan)
    a. *Niʔ   cən   txʷ-səplil.*
        AUX   1.SBJ   VBL-bread
        'I bought bread.'

    b. **Niʔ cən   txʷ-ət   kʼʷ   səplil.*
        AUX 1.SBJ buy-TR DET bread
        'I bought some bread.'

    c. *Niʔ   cən   ʔiləq-ət kʼʷ   səplil.*
        AUX 1.SBJ buy-TR DET bread
        'I bought some bread.'

Plus:

the incorporated noun must occur in a position different from the position of the argument when it is a full nominal, e.g. Turkish:

(17) Turkish (Aydemir 2004: 465-466)
    a. *Yasemin anahtar kaybet-ti.*   (**Anahtar Yasemin kaybetti.*)
        Yasemin key   lose-PST
        'Yasemin lost keys (or a key).'

    b. *Anahtar-ı Yasemin kaybet-ti.*
        key-ACC Yasemin lose-PST
        'Yasemin lost the key.' (OR: 'The key was lost by Yasemin.')

Thus, (17a) is not incorporation.

8. Serial verb constructions

A serial verb construction is a kind of "complex verbal predicate":

Definition 4: **Serial verb construction** (Haspelmath 2016: 296)
A serial verb construction is a monoclausal construction consisting of multiple independent verbs with no element linking them and with no predicate–argument relation between the verbs.

(18) a. Dagaare (Gur; Hiraiwa & Bodomo 2008: 796)
  ò   dà   sέ   lá   nέnὲ   ɔ̀ɔ̀
  3SG PST   roast FOC   meat   eat
  'He roasted meat and ate it.'

  b. Cantonese (Matthews 2006: 75)
  keoi⁵   haam³-sap¹-zo go   zam²tau⁴
  she   cry-wet-PFV   CLF   pillow
  'She made her pillow wet by crying.'

  c. Nêlêmwa (Oceanic; Bril 2004: 176)
  I   fuk   ulep   daxi   ni   fwaa-mwa.
  3SG   fly   cross.threshold up.away in   hole-house
  'It flies into the house.'

  d. Bislama (English-lexified creole; Crowley 2002: 223)
  Kali   i   katem splitem wud.
  Kali   3SG   cut   split   wood
  'Kali cut the log in two.'

requirements: – no linking element (distinguishing them from **converb** constructions)
  – independent verbs (unlike **auxiliaries**)
  – no predicate-argument relation between the two verbs (unlike **control** constructions)

Aikhenvald (2018: 1):
  "In many languages of the world, a sequence of several verbs **act together as one unit**. They form **one predicate**, and contain **no overt marker** of coordination, subordination, or syntactic dependency of any other sort. Such series of verbs are known as *serial verb constructions*, or serial verbs for short. Serial verbs describe what can be **conceptualized as a single event.** They are often pronounced as if they were **one word**."

But: – we do not know what "a single event" is
  – saying that they "act together as one unit" is too vague
  – the phonological criterion is not well-founded
  – control constructions are not excluded

Aikhenvald (2018: 18) criticizes Haspelmath (2016):

> "This definition is fairly problematic. ... The obscure formulation 'predicate–argument relation between the verbs' **arbitrarily excludes** serialization of **complement-clause taking predicates** and **causative serial verb constructions** (widespread types of serial verbs in many languages...). **Sharing tense, aspect, modality, and mood** are deemed to be 'unnecessary' criteria for serial verb constructions."

Aikhenvald is right that my 2016 definition is (partly) "arbitrary", like any other proposal for a definition – but it is motivated:

– complement and causative relations are not normally treated as SVCs
– one could add TAM sharing as a criterion, but it is unclear
   what such an additional criterion would exclude

(See also Lovestrand (2021) for recent discussion.)

Note that on my 2016 definition, **verb compounds are serial verb constructions**, e.g.

(5) Japanese
   John-ga   niwatori-o   naguri-korosi-ta.
   John-NOM chicken   beat-kill-PST
   'John beat a chicken and killed it.' (cited by Haspelmath 2016: 298)

9. Shared-core definitions of concepts for general linguistics

– several types of expressions that are typically treated as compounds are not covered by the present definitions.

– this is 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, e.g.

English Present Tense vs. German Present Tense:
   only the latter can regularly be used for future time reference as well
   (e.g. *wir kommen morgen* 'we will come tomorrow')

Turkish Dative case vs. Russian Dative case:
   only the latter case can be used for spatial goals as well
   (e.g. *Moskova'ya* 'to Moscow')

French Feminine gender class vs. English Feminine gender class:
   only the letter includes many inanimate nouns
   (e.g. *la lune* 'the moon'),

– similarly, German Compounds and Modern Greek 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.

10. Conclusion

– 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,
  – without having 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 "word-phrase" distinction or a "syntax-morphology" distinction

– is this progress? There is no empirical contribution in this paper!

– the definitions suggested here show that **there is no reason to be resigned** to a situation where we do not have definitions shared by the discipline.

– linguists often think of these terms as denoting **aspects of nature**, rather than **concepts created by linguists** in order to compare languages.

– they 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

BUT: if general terms of linguistics are **comparative concepts** rather than **natural kinds** (Haspelmath 2018), there is no reason to wait for better theories.

– there is no tendency for linguists' general-theoretical and methodological views and preferences to converge

  – so it is worth exploring the possibility of finding definitions that could work for everyone because **they do not rely on controversial theoretical views**

References

Aydemir, Yasemin. 2004. Are Turkish preverbal bare nouns syntactic arguments? *Linguistic Inquiry* 35(3). 465–474.

Bauer, Laurie. 2017. *Compounds and compounding* (Cambridge Studies in Linguistics). Cambridge: Cambridge University Press.

Bril, Isabelle. 2004. Complex nuclei in Oceanic languages: Contribution to an areal typology. In Isabelle Bril & Françoise Ozanne-Rivierre (eds.), *Complex predicates in Oceanic languages: Studies in the dynamics of binding and boundness*, 1–48. Berlin: Mouton de Gruyter.

Cetnarowska, Bożena. 2019. *Compound nouns and phrasal nouns in English and Polish* (Prace Naukowe Uniwersytetu Śląskiego w Katowicach). Katowice: Wydawnictwo Uniwersytetu Śląskiego.

Crowley, Terry. 2002. *Serial verbs in Oceanic: A descriptive typology*. 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 inalianability*, 65–110. Berlin: Mouton de Gruyter.

Gebhardt, Lewis. 2023. *The study of words: An introduction*. London: Routledge. (doi:10.4324/9781003030188)

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)

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. 2024. Inflection and derivation as traditional comparative concepts. *Linguistics (to appear)*.

Hiraiwa, Ken & Adams Bodomo. 2008. Object-sharing as symmetric sharing: Predicate clefting and serial verbs in Dàgáárè. *Natural Language & Linguistic Theory* 26(4). 795–832.

Lovestrand, 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)

Massam, Diane. 2017. Incorporation and pseudo-incorporation in syntax. *Oxford Research Encyclopedia of Linguistics*. (doi:10.1093/acrefore/9780199384655.013.190)

Matthews, Stephen. 2006. On serial verb constructions in Cantonese. In Alexandra Y. Aikhenvald & R. M. W. Dixon (eds.), *Serial verb constructions: A cross-linguistic typology*, 69–87. Oxford: Oxford University Press.

Mithun, Marianne. 1984. The evolution of noun incorporation. *Language* 60(4). 847–894.

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 & Mattiola, 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.

Schlücker, Barbara & Plag, Ingo. 2011. Compound or phrase? Analogy in naming. *Lingua* 121(9). 1539–1551. (doi:10.1016/j.lingua.2011.04.005)

Schlücker, Barbara. 2023. Compounding and linking elements in Germanic. *Oxford Research Encyclopedia of Linguistics* 2023. (doi:10.1093/acrefore/9780199384655.013.954)

Trips, Carola & Kornfilt, Jaklin. 2015. Typological aspects of phrasal compounds in English, German, Turkish and Turkic. *STUF - Language Typology and Universals*. De Gruyter (A) 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.