CHAPTER 2

WORD-CLASS UNIVERSALS AND LANGUAGE-PARTICULAR ANALYSIS

MARTIN HASPELMATH

2.1 THREE WORD-CLASS UNIVERSALS

THIS chapter considers both how word classes are analysed or described in particular languages (section 2.3), and what we can say in general about major word classes. We will see that these two questions are less directly related than is often thought. We begin with some important word-class universals in this first section, as well as some cross-linguistic macro-types of word-class patterns in section 2.2. Word-class universals are not particularly well known, although many comparative linguists are aware of them and will not be surprised by the Universals 1–3 given below, even if they have not read Croft (1991; 2000).

Much of the comparative literature on word classes asks to what extent the distinction between nouns, verbs, and adjectives is universal (e.g. Sasse 1993; Evans 2000; Baker 2003; Dixon 2010: ch. 11). This question (which I refer to as the 'distinctness question') has no clear answer, so it will be discussed only later (section 2.4), and I first focus on universals (section 2.1) and cross-linguistic macro-types (section 2.2).

Universals 1–3 (due to Croft 1991: ch. 3) are about the occurrence of function indicators, i.e. copulas, attributivizers, and termifiers, in three different propositional act functions: predication, modification, and reference, with three different semantic root classes: action, property, and object roots. All these terms will be explained and discussed further below, and the universals will become clearer once they are exemplified.

> **Universal 1**
> If a language has a copula, i.e. a special form that indicates predicative function, it is used with object roots and/or property roots.
>
> **Universal 2**
> If a language has an attributivizer, i.e. a special form that indicates modifying function, it is used with action roots and/or object roots.
>
> **Universal 3**
> If a language has a termifier (= nominalizer or substantivizer), i.e. a special form that indicates referential function, it is used with property roots and/or action roots.

The term triple 'predication/modification/reference' for the propositional act functions is familiar to every linguist, but of the three types of function indicators, only the first has a well-known existing term (*copula*). But the main subtypes of the other function indicators are also well known: relativizers and genitive flags (the two subtypes of attributivizers seen in (3a) and (4a)), as well as action nominalizers and substantivizers (the two subtypes of termifiers seen in (5a) and (6a)).

The reason the universals are formulated in terms of roots, not in terms of 'words', is that there is no general cross-linguistic understanding of what a 'word' is (other than as defined by the conventional orthographic representation). But in many or most languages, many roots can be words (especially nouns like *dog* and adjectives like *big*), and whenever a root cannot occur on its own without an affix (e.g. when tense affixes or person indexes are obligatory in verbs), these obligatory elements are not immediately relevant to Universals 1–3. Thus, formulating the universals in this way makes them generally applicable, and we do not need to worry about 'words'.

The construction types that the three universals make claims about are summarized in Table 2.1, where the boldfaced elements in the examples are the function indicators.

Table 2.1 Three propositional act functions and three semantic root classes

	reference: marked by termifiers	modification: marked by attributivizers	predication: marked by copulas
objects	—	genitive flag ((4a): *the rent of the house*)	object–word copula ((1a): ***is** a student*)
properties	substantivizer ((5a): *the new **one***)	—	property copula ((2a): ***is** big*)
actions	nominalizer ((6a): *the open-**ing***)	relativizer (3a: *the work **that** they did*)	—

What the function indicators have in common is that they indicate an unusual (unexpected, surprising) propositional act function: a genitive flag indicates that an object root (unexpectedly, surprisingly) has modifying function, a property copula indicates that a property root has predicating function, a nominalizer indicates that an action root (unusually) has referential function,[1] and so on.

After this short overview, let us now consider each of the universals individually. For convenience, the Universals 1–3 are repeated in what follows.

> **Universal 1**
> If a language has a copula, i.e. a special form that indicates predicative function, it is used with object roots or property roots.

So, we find many languages which require a copula with object–word predicates as in (1a), though not all languages have them (1b). Likewise, many languages require a copula with property predicates (2a), but some have no copula here (2b).

---

[1] For action nominalization constructions, see Chapter 5 in this volume.

(1)   a.   Italian (copula *è*)
             Michele   è   studente.
             Michele   is   student
             'Michele is a student.'

       b.   Russian (no copula)
            Миша—студент.
            Miša   student.
            Misha   student
            'Misha is a student.'

(2)   a.   English (copula *is*)
            *Our dog **is** big.*

       b.   Mandarin Chinese (no copula)
            我們的狗很大
            Wǒmen   de   gǒu   hěn   dà.
            we   GEN   dog   very   big
            'Our dog is big.'

While a function indicator is commonly used for predicating object or property roots, it is not always necessary, so Universal 1 must be formulated as an implicational universal. But, crucially, action roots generally do not require a form that indicates predicative function (in fact, they probably never do, as I do not know of a single example).

Universal 2 is analogous to Universal 1:

Universal 2
If a language has an attributivizer, i.e. a special form that indicates modifying function, it is used with action roots or object roots.

The two main types of attributivizers are relativizers (relative markers, including participial affixes) and genitive flags (affixes or adpositions). So, we find many languages which require a relativizer with action modifiers (= relative clauses) as in (3a), though not all languages have them (3b). Likewise, many languages require a genitive flag with object–word modifiers (4a), but some have no marker here (4b).

(3)   a.   Lezgian (relativizer *-j*) (Haspelmath 1993a: 344)
            *awu-nwa-**j**   k'walax-ar*
            [do-PRF-REL]   work-PL
            'the work that had been done'

       b.   Japanese (no relativizer)
            私がした仕事
            watashi   ga   shi-ta   shigoto
            [I   NOM   do-PST]   work
            'the work I did'

(4)   a.   Cape Verdean Creole (genitive flag *di+*) (Baptista 2013)
           *kel   renda   **di**   kaza*
           the   rent   of   house
           'the rent of the house'

      b.   Seychelles Creole (Michaelis & Rosalie 2013)
           lakaz   sa   zonm
           house   that   man
           'that man's house'

So, again, the function indicators do not occur universally, but property words are the least likely to require an attributivizer.

Finally, the situation is again completely analogous with indicators of referential function:

> **Universal 3**
> If a language has a termifier (= nominalizer or substantivizer), i.e. a special form that indicates referential function, it is used with property roots or action roots.

The two main types of termifiers are substantivizers and nominalizers.[2] So we find some languages which require a substantivizer with property-denoting referential expressions as in (5a), though not all languages have them (5b). Likewise, many languages require a nominalizer with action-denoting referential expressions (6a), but some have no marker here (6b).

(5)   a.   English (substantivizer +one)
           *the new **one***

      b.   Spanish (no substantivizer)
           la   nueva
           the   new
           'the new one'

(6)   a.   German (nominalizer *-ung*)
           *die Öffn-**ung** des Fensters*
           the open-NMLZ of.the window
           'the opening of the window'

      b.   Malay (no nominalizer) (Yap et al. 2011: 13)
           Makan   lewat   tak   bagus.
           eat   late   not   good
           'Eating late is not good.'

---

[2] The term *termifier* (created for the purposes of this chapter) is derived from *term*, a synonym of *nominal (expression)* (*term* is used, for example, by Dik (1997)). The term *termifier* is probably not really necessary outside of the current comparative context, but it nicely serves to highlight the parallel with attributivizers and copulas (= predicativizers). (As Eva van Lier points out (p.c.), a term such as *referentializer* would also be possible.)

In addition to substantivizers, property roots may also occur with abstract-noun markers, as in English *new-ness*, or Spanish *nov-edad*. Such markers are semantically similar to action nominalizers, as the resulting form refers to the property as such, not to an entity bearing the property. Such markers are left aside in the present chapter for expository reasons, but the main generalizations should apply to them, too.

In the world's languages, the three pairings object–reference, property–modification, and action–predication have no special function indicators corresponding to them (though this would be logically possible). Table 2.1 thus shows dashes for these pairings. Croft (1991) describes these as 'unmarked combinations', and he shows that the lack of special marking is due to frequency differences: The usual pairings are much more frequent in discourse than the unusual pairings (there is thus a kind of form–frequency correspondence here).

Now, crucially, it is logically possible that these pairings might have function indicators as well, so there might be languages that have copulas with all three root classes (as in (7)), attributivizers with all three root classes (as in (8)), and termifiers with all three root classes (as in (9)).

(7) hypothetical all-copula language (copula BE)
    a. Michael BE student.
    b. Dog BE big.
    c. **Penny BE run home.**   **(action–predication)**

(8) hypothetical all-attributivizer language (attributivizer OF)
    a. house OF man
    b. **house OF new**   **(property–modification)**
    c. house OF I bought

(9) hypothetical all-referentializer language (termifier ONE)
    a. **I saw the teacher ONE.**   **(object–reference)**
    b. I saw the new ONE.
    c. I saw Penny's run ONE home.   ('I saw Penny running home.')

The claim of Universals 1–3 is that if a language has such an unlikely function indicator, it will also have the more likely ones.

This may not sound like a very strong claim, but it needs to be formulated in this somewhat weak way because there are many constructions where no function indicators are used, i.e. where the coding is not asymmetric: the (b) cases in (1)–(6). In addition to such cases of symmetric zero coding, as well as symmetric overt coding (as in (7)–(9), and as sometimes actually attested, e.g. in (20)), it is logically also possible to have **counter-asymmetric coding**: a function indicator only in the usual pairings object–reference, property–modification, and action–predication. The latter is what the Universals 1–3 exclude, and this tendency for asymmetric coding can be explained as due to frequency-induced predictability and efficient coding (Haspelmath 2021).

So far, I have not used the terms *noun*, *verb*, or *adjective*. This is deliberate, because the use of these terms in general contexts has led to a lot of confusion in the past (see also Croft 2022: ch. 2). It is worth emphasizing that we do not need them for stating these key observations about grammatical coding and their efficiency-based explanation. In the next section, we will see what role the traditional word-class labels can play in a cross-linguistic context.

2.2 MACRO-TYPES: INDICATOR COEXPRESSION PATTERNS

2.2.1 Alignment types and indicator coexpression types

The word-class universals of section 2.1 are similar to the universals of argument marking (flagging and indexing; see Haspelmath 2005; Dryer 2007; Siewierska & Bakker 2009; Bickel 2011).[3] In order to highlight the major types, linguists have set up a tripartite semantic map with different coexpression ('alignment') patterns, as shown in Figure 2.1.
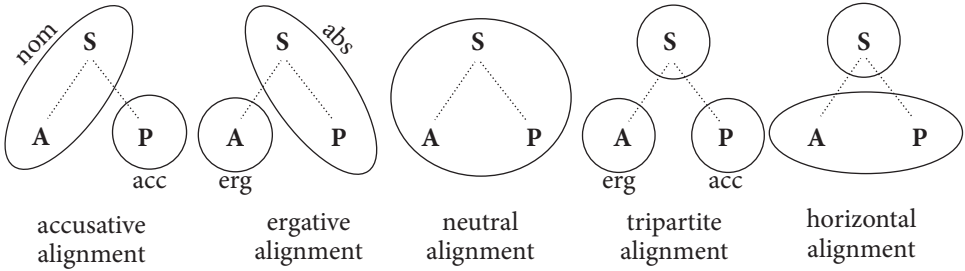

FIGURE 2.1 Five alignment types of argument markers

Labels for the two kinds of argument marker (flags and index sets) usually have terms that are taken from these alignment patterns (accusative case, ergative case, absolutive indexing, etc.), and as shown in Haspelmath (2005; 2015), a very similar approach works well for ditransitive alignment.

For word-class coding patterns, the three elements on the semantic map are the three root classes object (*ob*), action (*ac*), and property (*py*).[4] Figure 2.2 is parallel to Figure 2.1.

---

[3] That the comparison of word classes involves very similar issues as the comparison of syntactic role classes ('grammatical relations') was made particularly clear by Dryer (1997: §2).

[4] The terminology in the literature varies a little; in Haspelmath (2012a), I used *thing* for *object*, and some authors use *event* for *action*. There is no difference in substance here, though actions are generally thought to be subtypes of events. The universals are unaffected by this difference. (The reason I used *thing* instead of *object* was that the latter term is also used prominently for a syntactic role type (*subject*, *object*, *oblique*); but here I use *object*, to conform with Croft's usage.)
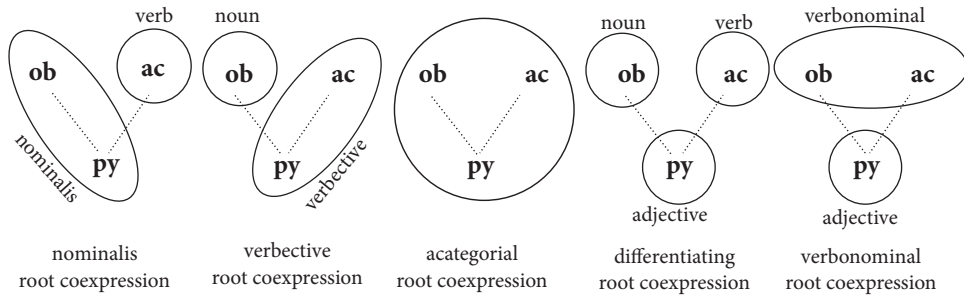
FIGURE 2.2 Five coexpression types of function indicators

The three terms *nominalis*, *verbective*, and *verbonominal* for three of the coexpression patterns are new and perhaps surprising.[5] By contrast, the terms *verb*, *noun*, and *adjective* for non-coexpressing parts of the maps are quite familiar. I did not need them in the statement of the universals in section 2.1, but here we are talking about the ways in which the semantic root classes differ in their coding, so it is here that the old grammatical terms come to play a role (see also section 2.5).

Just as different aspects of argument coding (flagging vs indexing) can have different alignment types (e.g. ergative case marking but accusative indexing), we may have different indicator coexpression patterns for different aspects of coding.[6] In the following sections, we will see examples of the five coexpression types in the three propositional act functions: predicative (section 2.2.2), referential (section 2.2.3), and modifying function (section 2.2.4).

2.2.2 Coexpression types in predicative function

In (10), I give an overview of the five predicative coexpression types, with a few initial examples. They are illustrated further below.[7]

(10)   a. nominalis coexpression
        e.g. a nominalis copula, as in English (*be* for both noun and adjective)

---

[5] Alternatively, one could use the terms *anti-verb*, *anti-noun, and anti-adjective* (David Gil (p.c.)). They would work quite well, too, and are transparent in the context of the semantic map in Figure 2.2 (though less so if one is not aware of this context).

[6] In alignment patterns of argument markers, the tripartite and horizontal types are very rare because they make unnecessary distinctions (S never cooccurs in the same clause with A or P, so it can be conflated with either or both of them). This motivating factor is absent in the coexpression patterns of function indicators, so there is no a priori reason to expect that only three of the five logically possible patterns are attested.

[7] The second line in (10a–e) is preceded by 'e.g.' because, in each case, there are alternative logical possibilities for the coexpression patterns. However, due to the frequent absence of function indicators in the usual ('unmarked') pairings (as seen in Universals 1–3), the types mentioned here are de facto the only ones.

    b. verbective coexpression
       e.g. a noun copula, as in Mandarin Chinese (是*shì*)

    c. acategorial coexpression
       e.g. no copula at all, as in Sri Lanka Malay (ex. 12) and Lillooet (ex. 13)

    d. differentiating pattern (no coexpression)
       e.g. different copulas for nouns and adjectives, as in Buwal (ex. 14)

    e. verbonominal coexpression
       e.g. copula only for adjectives, but not for nouns (nonexistent?)

The NOMINALIS type has this label because in the earlier Western tradition, the Latin terms *nomen* and *nominalis* were used both for nouns and adjectives (these were regarded as subtypes of the *nomen* class, called *nomen substantivum* and *nomen adjectivum*; see Chapter 27 in this volume). This coexpression type is well known from Indo-European languages like Latin and English, but also occurs elsewhere (Stassen 1997: xxx). It is so well known that no further illustration is needed (see (1a) and (2a)).

The VERBECTIVE type, where event predication and property predication are treated in the same way, is also well known, and is found, for example, in Cantonese (Francis & Matthews 2005: 274), which has an obligatory copula *hai* 'be' for object–word predicates (corresponding to Mandarin *shì*), but no copula with action and property predicates:

(11)   Cantonese
      a. *M   hai   hoksaang.*
         not   be   student
         'He's not a student.'

      b. *M   haam.*
         not   cry
         'He's not crying.'

      c. *M   hausaang.*
         not   young
         'He's not young.'

The ACATEGORIAL type is not so uncommon in predicative position, because there are many languages that lack a copula. In Sri Lanka Malay (Nordhoff 2013: §9.3), for example, there is no nominal copula (12a), and property roots (12b) look even more like action roots (12c) in that they can also be preceded by tense-aspect markers such as *arà-*:

(12)   Sri Lanka Malay (Nordhoff 2013: 252, 255, 250)
      a. *Sindbad   hatthu   Muslim.*
         Sindbad   INDEF   Muslim
         'Sindbad was a Moor.'

    b. *Ruuma   arà-kiccil.*
       house   NPST-small
       'The houses are getting small.'

    c. *Incayang   arà-maakang.*
       3SG   NPST-eat
       'He is eating.'

Another example of the acategorial type is Lillooet (a Salishan language), where predicative elements are always clause-initial, again without any copula (Davis et al. 2014: 196).

(13)   Lillooet
    a. *Šmúłač   ta=kʷúkʷpiʔa.*
       woman   DET=chief
       'The chief is a woman.'

    b. *Láχləχ   ta=kʷúkʷpiʔa.*
       smart   DET=chief
       'The chief is smart.'

    c. *λ'iq   ta=kʷúkʷpiʔa.*
       arrive   DET=chief
       'The chief arrived.'

A similar pattern is found in Southern Wakashan languages such as Nuu-chah-nulth (discussed in Chapter 6 in this volume). For more on Salishan languages, see Chapter 29 in this volume.

    The DIFFERENTIATING type would be represented by languages that have two different copulas, one for object roots (nouns) and one for property roots (adjectives). They seem to be quite rare, but the Chadic language Buwal is described in this way by Viljoen (2013): the nominal copula is *ārā*, while the adjectival copula is *ndzā*.

(14)   Buwal (Viljoen 2013: 448, 187)
    a. *mbàw   ārā   dādāwār*
       child   COP   evil.person
       'The child is an evil person.'

    b. *ā-ndzā   ɓārɓār*
       3SG.SBJ-COP   hard
       'It is hard.'

    Finally, the VERBONOMINAL type would be represented by a language that only has a copula for property words, but such languages do not seem to exist: Stassen (1997) and Pustet (2003) note that if a language has a copula for property word predication, it also has a copula for object word predication.

2.2.3   Coexpression types in referential function

Moving on to the second propositional act function, (15) gives an overview of the coexpression types in referring function, with a few initial examples.

(15)   a. nominalis coexpression
          e.g. no substantivizer, as in Spanish (*la casa*, *la nueva* (ex. 5b))

       b. verbective coexpression
          e.g. identical termifier for properties and actions (nonexistent?)

       c. acategorial coexpression
          e.g. no termifier, as in Tagalog ((ex. 16))

       d. differentiating pattern (no coexpression)
          e.g. different substantivizer and nominalizer, as in English
          (*the house*, *the new **one***, *the open-**ing***)

       e. verbonominal coexpression: unattested
          e.g. substantivizer but no nominalizer (nonexistent?)

The NOMINALIS coexpression type is again the type found in the traditional Indo-European languages such as Latin, and also in Spanish and German. Both object roots and property roots (but not action roots) can be used in referring function with an article in these languages, e.g. Spanish *la casa* 'the house', *la nueva* 'the new one' (see (5b); German *das Haus* 'the house', *das neue* 'the new one'. By contrast, action roots in referring function need some kind of relativizer or nominalizer.

The VERBECTIVE type would be represented by a language which uses a general termifier, i.e. the same marker as a substantivizer ('the new-TERM' = the new one) and as a nominalizer ('the open-TERM' = the open-ing). Such languages do not seem to exist.

The ACATEGORIAL type is famously found in Classical Nahuatl and also in Tagalog, as illustrated by the examples in (16) from Gil (1993). Not only object roots as in (16a) and property roots as in (16b), but also action roots as in (16c) can be used in referring function following a role-marking proclitic (in these examples, it is always the Nominative marker *ang*).

(16)   Tagalog (Gil 1993: 1140)
       a. *Lumabas   ang   bangkero.*
          went.out   NOM   boatman
          'The boatman went out.'

       b. *Lumabas   ang   mabait.*
          went.out   NOM   kind
          'The kind one went out.'

       c. *Lumabas   ang   bumalik.*
          went.out   NOM   returned
          'The one who returned went out.'

The claim here is that the role marker is not a termifier, and if we say the same about the Lillooet determiner *ta* = in (13), then Lillooet also exemplifies this type (because Lillooet allows reversing all of (13a–c): *kʷúkʷpíʔa ta-lə́χləχ* 'the smart one is a chief', etc.).

Next, the DIFFERENTIATING type is represented by English, which has two different termifiers: a substantivizer *one* for property roots (*the new **one***), and an action nominalizer *-ing* (in *the open-**ing***).

Finally, the VERBONOMINAL type would be represented by languages where adjectives need a substantivizer, but verbs do not need a nominalizer. Again, such languages may not exist.

Now before we get to the coexpression types in modifying function in the next section, let us consider the first two propositional act functions together. There are two clear types here: the DOUBLY NOMINALIS type represented by Spanish (coexpression of objects and properties in both functions), and the DOUBLY ACATEGORIAL type represented by Lillooet and Tagalog. These are the kinds of languages that have sometimes been said to lack a noun–verb(–adjective) distinction, though in different ways: in languages like Spanish (and especially in Latin), nouns and verbs have been said to be subclasses of a larger 'nomen' class, also because they have very similar number marking (and in Latin case marking) properties.[8] In Lillooet and Tagalog, the reason for saying that they are acategorial has been that they lack a copula in predicative function. The doubly DIFFERENTIATING TYPE may exist as well, but is much less prominent.

If one focuses one's attention on the two salient types, one may come to the conclusion that Spanish/Latin-type languages have a single ('flexible', or 'merged') noun–adjective class, while languages like Lillooet and Tagalog only have a single noun–verb–adjective class. There is nothing wrong with this, in principle, because one may choose one's types as one pleases, and, at first glance, it seems interesting that the doubly nominalis Spanish type seems to be fairly common, while the doubly acategorial type of Tagalog and Lillooet seems to be quite rare.

But we need to remember that we know nothing about correlations at this point: is the doubly nominalis type more common than expected by chance (i.e. expected on the basis of the combined probabilities of nominalis predication and nominalis referential use)? Is the doubly differentiating type less common than expected by chance? For the doubly acategorial type, we can perhaps exclude this possibility, because acategorial (= copulaless) predication is not uncommon (see Sri Lanka Malay in (12)), whereas acategorial referential use of the Tagalog or Lillooet type is quite uncommon. But we would need more systematic cross-linguistic data before quantitative statements can be made.

Thus, it is too early to make generalizations about 'Spanish-type' languages, let alone 'Tagalog-type' languages. We do not know how common these language types are, and whether these types are significant. Moreover, we should not neglect the modifying function, which we will consider in the next section.

---

[8] The fact that Spanish and Latin use the same copula in predicative function has been less salient in the traditional discussion, but if they used two different copulas (like Buwal in (14)), they would probably not have been lumped together by anyone.

2.2.4 Coexpression types in modifying function

Finally, (17) gives an overview of the coexpression types in modifying function.

(17) a. nominalis coexpression
          e.g. relativizer for action modification vs uncoded adjective and
          uncoded possessor, as in Seychelles Creole (ex. 18)

      b. verbective coexpression
          e.g. identical attributivizer for action and property modification,
          as in Archi (ex. 19)

      c. acategorial coexpression
          e.g. a general attributivizer, as in Mandarin Chinese *de* (ex. 20)

      d. differentiating pattern (no coexpression)
          e.g. different genitive flag and relativizer, as in English (genitive flag *of*
          vs relativizer *that*)

      e. verbonominal coexpression
          e.g. identical relativizer and genitive flag, as in Aramaic (ex. 21)

The NOMINALIS type is found, for example, in Seychelles Creole, where there is no marker for possessive object–word modifiers (18a) or property modifiers (18b), whereas action modifiers (relative clauses) have a relativizer *ki* (18c) (Michaelis & Rosalie 2013).

(18) Seychelles Creole
      a. *lakaz   Marcel*
          house   Marcel
          'Marcel's house' (see (4b))

      b. *dilo   so*
          water   hot
          'hot water'

      c. *sa   zoli   lakaz   ki   ou   annan   la*
          DEM   nice   house   REL   you   have   there
          'this nice house that you have there'

The VERBECTIVE type is found, for example, in Archi, which has an attributive suffix *-t:u* for action and property modifiers, but a Genitive suffix *-n* (Chumakina 2018). Some examples of the attributivizer are given in (19).

(19) Archi (Chumakina 2018: 177)
      a. *mu-t:u-b   noʕš*
          beautiful-ATTR-G3   horse(G3)
          'beautiful horse'

b. *kwaršu-t:u-b   χabar*
happen-ATTR-G3   story(G3)
'story that happened'

This type seems to be widely represented among the world's languages. The *WALS* chapter by Gil (2005) gives 33 languages from all continents that 'collapse adjectives and relative clauses'.

The ACATEGORIAL type is represented, for example, by Mandarin Chinese, which has an attributivizer *de* that is used with object roots, property roots, and action roots. For example:

(20) a. 父親的房子
fùqīn   de   fángzi
father   ATT   house
'father's house'

b. 漂亮的房子
piàoliang   de   fángzi
beautiful   ATT   house
'beautiful house'

c. 笑的孩子們
xiào   de   háizi-men
laugh   ATT   child-PL
'laughing children'

The DIFFERENTIATING type is again represented by English, which has two genitive flags (postclitic *'s*, preposition *of*) for object–word modifiers, no marking for property modifiers, and relativizers (*that*, *-ing*) for action modifiers.

Finally, the VERBONOMINAL type is represented by those languages that have a relativizer with the same shape as the genitive marker, which is not used with property words. Gil (2005) finds this type only in two languages, one of which is a variety of Aramaic. Fassberg (2019) describes Western Aramaic, where a particle *ti* can be used both as a genitive flag and as a relativizer.

(21) Western Aramaic (Fassberg 2019: 648)
a. *so:ba   ti   blo:ta*
mayor   of   village
'the mayor of the village'

b. *hanna   γamla   ti   tˤʕil-le*
this   camel   that   it.carried-him
'this camel which carried him'

2.2.5 Coexpression patterns across the three propositional act functions

If we now continue to compare the patterns across the three functions, we do not find much further evidence for a dominant coexpression pattern of function indicators. Recall that we suggested that 'Spanish-type' languages might have a general nominalis type, but at least in Latin, Spanish, and German, this does not extend to the modifying function (which is of the differentiating type, as in English).

Perhaps most strikingly, the acategorial type does not extend even to Tagalog (see Chapter 35 in this volume). Object–word modifiers are indicated by a genitive flag *nang* (22a), while property and action modifiers are indicated by an attributivizer *na/ng* (22b–c).

(22)   a.   *ang   bahay   n(an)g   ama*
            NOM   house   GEN   father
            'father's house'

       b.   *ang   maliit   na   bahay*
            NOM   small   ATT   house
            'small house'

       c.   *ang   babae=ng   nagbabasa   n(an)g   diyaryo*
            NOM   woman=ATT   read.IPV   GEN   newspaper
            'the woman who is reading a newspaper'

Thus, as far as we know at this moment, there is no strong tendency for coexpression patterns to cluster beyond the individual propositional act functions. While linguists generally have the feeling that there is something like a 'dominant role alignment' across argument-marking constructions (and we often even say that a language is 'an ergative language'), there is no good evidence, at this point, for a notion of 'dominant coexpression pattern' across propositional act functions. Thus, from the point of view that is adopted in this section, we cannot say (so far) that in general, languages have 'flexible' or 'differentiated' root classes (see Rijkhoff & van Lier 2013).

2.3 LANGUAGE-PARTICULAR ANALYSIS CREATES CONSTRUCTION-BASED CLASSES

There has long been broad agreement that languages differ in interesting ways in the way they group words into word classes (or *parts of speech*, or *lexical categories*). Sapir (1921: 118) wrote: 'Each language has its own scheme [of parts of speech]. Everything depends on the formal demarcations which it recognizes.' And Schachter & Shopen (2007: 1) wrote: 'There are striking differences between languages with respect to both the kind and the number of distinctions.'[9]

---

[9] See also Himmelmann (2008: 259): 'Lexical and syntactic categories are by definition language-specific as they are based on language-specific formal features and the distribution of such features tends to show language-specific idiosyncrasies.'

But what exactly do these authors mean by 'demarcations' or 'distinctions'? This is not immediately clear, because there is a wide variety of ways in which the roots, grammatical markers, and complex expressions are grouped into morphosyntactically relevant classes.

For example, about one-third of verbs and adjectives in Godoberi (a Nakh–Dagestanian language) have an initial gender marker, while two-thirds lack such a marker (Kibrik 1996: 24, 44). So one could set up a word-class 'Gendered Verbective' and another word-class 'Genderless Verbective'.[10] Similarly, for German, one could set up a word-class 'Very-Modifiable' for those verbs and adjectives that can be modified by *sehr* 'very' (e.g. degree adjectives such as *schön*, and 'degree verbs' such as *sich freuen* 'be happy'; but not verbs like *arbeiten* 'work'). Or for Maricopa (a Yuman language), one could distinguish between Pluralizable words (most verbs and some nouns) and Nonpluralizable words (all the others; Gordon 1986: 29, 90).

These examples may appear surprising, but that we can bring them up here simply follows from the logic of grammatical analysis: we formulate general (nonsemantic) regularities that govern the cooccurrence of classes of forms. These nonsemantic rules are called CONSTRUCTIONS,[11] and the classes of forms are FORM CLASSES (Bloomfield 1933: 146). The identification of form-classes is sometimes called distributional analysis (see section 11.2 in this volume).[12] Such form-classes are often cross-classified, as in Godoberi, which has four classes of verbs/adjectives:

  (i) Gendered Tense-marked (e.g. *ičã*: 'to sell')
  (ii) Gendered Tenseless (e.g. *mik'isi*: 'young')
  (iii) Genderless Tense-marked (e.g. *bit'i*: 'to tear')
  (iv) Genderless Tenseless (e.g. *q'aruma*: 'greedy')

Which of the classes (Gendered/Genderless or Tense-marked/Tenseless) is taken as the superclass here and which is the subclass is essentially arbitrary, and is determined largely by pedagogical considerations. For reasons of grammatical tradition, linguists will almost always give preference to the labels 'noun', 'verb', and 'adjective' (as was done at the end of the preceding paragraph, and as is done in Kibrik (1996)), but this does not mean that these classes are in fact privileged in the languages.

Many linguists have the feeling that when we consider all the relevant facts, we find again and again that the most important (or privileged) form classes are noun, verb, and adjective, and it may be that this will eventually turn out to be the case. But, in practice, we never consider all the facts—not when we describe the word classes of a single language, and not at all when we compare languages, because the various language-particular constructions are so hard to compare. So de facto, we tend to look for classes that resemble our stereotypical nouns, verbs, and adjectives, and often we find some 'criteria' for identifying classes with these well-known labels in particular languages. But language analysis (= description) is not

---

[10] Recall that 'verbective' as introduced above as a term for classes that neutralize the property vs action distinction.

[11] By *nonsemantic*, I simply mean regularities of cooccurrence that do not follow from semantic (in)compatibility. Thus, #*I will leave yesterday* is semantically ill-formed, but **I will tomorrow leave* is syntactically ill-formed.

[12] The term *distributional analysis* may seem to suggest a specific approach, but in fact this is the basis of all morphological and syntactic analysis. There is no reason to associate it with a particular theoretical or ideological orientation, because it is universally adopted in linguistics.

about finding criteria to identify pre-established classes[13]—it is about creating those classes that we need to describe the constructions of a language.[14] And these classes are very numerous once one considers all the facts (probably many hundreds of classes for all languages; this is also observed by Croft in this volume—see section 11.6).

In reality, we do not know which kinds of classes are 'privileged', either in particular languages or in general. The notion of privileged classes may not mean much for language-particular analysis, where we need a complete picture anyway, and cannot be content with a few major classes. So, apart from pedagogical considerations, it does not really matter how exactly we set up the major classes and their subclasses. And, for comparative purposes, all we can do is focus on particular kinds of differences that happen to be readily comparable, as we did in sections 2.1–2.2. Intuitively, it seems that the distribution of function indicators is the most important criterion for nouns/verbs/adjectives across languages (but see section 2.7 on other markers).

In addition to 'major classes' (for which we use the labels *noun*, *verb*, *adjective*), all languages also have forms that do not readily fit into such larger classes, e.g. pronouns, numerals, quantifiers, demonstratives, coordinators, subordinators, as well as adverbs and particles. In the older 'parts of speech' tradition (see Chapter 27 in this volume), seven of these were included as specific parts of speech (numerals, interjections, adverbs, prepositions, conjunctions, pronouns, articles), but here the degree of cross-classification and arbitrariness of decisions is even greater than in the case of nouns, verbs, and adjectives. The reason why these old and plainly unsatisfactory terms are still around is that there is no alternative general set of classes. As Sapir noted: each language has its own classes, and what is general across languages is not kinds of classes, but universals like those in section 2.1.

2.4 WHY THE DISTINCTNESS QUESTION CANNOT BE ANSWERED

In the past, many linguists have asked whether two semantic classes (of the type object, action, property) are 'distinct' or not, either in a particular language or in the world's languages in general. We often find questions such as those quoted in (23).

(23)   a. Robins (1952: 296): 'Are we then able to say that there are any universal categories in grammar other than the purely segmental ones?'

---

[13] It is therefore surprising how much effort the general literature tends to invest in 'finding criteria' (e.g. Sasse 1993; Beck 2002), as if we already knew which classes there are and we only had to match these pre-existing classes to language-particular phenomena. But the point of the comments at the beginning of this section was precisely to remind us that there are no pre-existing classes.

[14] It is thus wrong to ask, in a questionnaire for descriptive grammars, whether there are 'operational definitions' of the classes noun, verb, adjective, etc. (Comrie & Smith 1977: §1.16). The operational definitions must be language-particular, and they cannot define general categories.

b. Evans (2000: 720): 'Are there languages which go further [than merging verbs and adjectives], merging nouns and verbs into a single class of predicates?'

c. Chung (2012): 'Are lexical categories universal?' (title of paper)

d. Davis et al. (2014: e195): 'The empirical question addressed in our second case study is: "Do all languages have a distinction between nouns and verbs?"'

e. Sasse (2015: 166): 'The articles in Vogel & Comrie (2000) provide a good overview of the more recent landscape of cross-linguistic word class research and its central controversies and proposals. The most fundamental point in dispute … is the universality of word classes.'

However, these questions cannot be answered when there are no limits on what can count as a 'distinction' between classes and when at the same time the classes are not defined independently of the language-particular grammatical constructions (the same point is made by Croft in this volume—see section 11.3). The latter point is of course universally recognized: word classes (of particular languages) are defined in terms of grammatical constructions, not in terms of 'notional categories'. The 'traditional notional classes' do not necessarily correspond to grammatical classes in particular languages.[15]

So if word classes are defined in a language-particular way, with reference to different constructions in different languages (as seen in section 2.3), then there is no way to match classes across languages. Rigorous comparative grammar requires uniformly defined yardsticks for 'measurement' (comparative concepts—see Haspelmath 2018), so it is not admissible to apply different criteria in different languages for identifying, say, adjectives or verbs. But this is precisely what Baker (2003) does (as described by Croft 2009: §3), and it is also what Chung (2012) does (as described by Haspelmath 2012b).[16]

The conventional way of hoping to resolve this problem is to say that the classes are set up by language-particular constructional (or distributional) criteria, but are matched across languages by their semantics:

> [With reference to 'formal', distributional criteria], we can decide for each word in the language to what syntactic class or classes it belongs. It is true that not all the members of class X [English *boy*, *woman*, *grass*, *atom*, *tree*, *cow*, *truth*, *beauty* …] denote persons, places and things … However, it may still be true that all (or the vast majority) of the lexical items which refer to persons, places and things fall within the class X; and, if this is so, we may call X the class of nouns.
>
> (Lyons 1968: 318)

---

[15] For example, Lyons (1968: 317): 'If the class of nouns is defined in "notional" terms, as that class of lexical items whose members denote places, persons and things … the definition cannot be applied without circularity to determine the status of such English words as *truth*, *beauty*, *electricity*, etc.'

[16] See also Croft & van Lier's (2012) critique of Chung (2012), and Haspelmath's (2014) critique of Davis et al. (2014).

For very similar statements, see Evans (2000: 709) and Schachter & Shopen (2007: 2). But this does not solve the problem, because it merely provides a justification for labelling. Of course, it would not be very reader-friendly to set up a class of nouns and to call it 'Class Y' (this was Garvin's (1951) approach for the North American language Kutenai), but it would highlight the fact that this class is not defined in the same way as the class that we call 'noun' in English. There is no theoretical reason to equate the two classes of words.

As we saw in section 2.3, there are many different ways in which word–size elements can be arranged in classes and subclasses. For example, one might say that English words like *Pat*, *Kim*, or *Lee* are a subclass of the English class 'Noun', or one might say that they constitute a separate class 'Proper Name' (though there would of course have to be a super-class comprising both Nouns and Proper Names). Which of the many different classes that must be created for the purpose of language-particular analysis are to be known by which label is quite arbitrary. This is recognized by some authors,[17] but many still treat it as a sub-stantive question (and they sometimes even claim that word classes are part of an innate grammar blueprint or 'universal grammar').[18]

That there is no substantive question here was clearly recognized by Croft (2000: 65): 'Noun, verb and adjective are not categories of particular languages'. What is general across languages is implicational universals of coding like those seen in sections 2.1–2.2, but not the classes noun, verb, and adjective, and it is not even clear what it would mean for these classes to be universal as grammatical classes.[19] Some cross-linguistic work on word classes has recognized this and has focused on finding generalizations about the various kinds of markers, rather than about the grammatical classes (e.g. Pustet 2003 on copulas; Ye 2021 on markers associated with property words/adjectives).

2.5 Nouns, verbs, and adjectives as concepts of general grammar

We have seen that we do not need the concepts of noun, verb, and adjective for stating universals and distinguishing macro-types (sections 2.1–2.2), and that different languages have a large number of classes that cannot be mapped onto each other because they are based on different language-particular constructions (section 2.3). Such classes can (for mnemonic

[17] Schachter & Shopen (2007: 13), after discussing Tagalog and Nuu-chah-nulth word classes, con-clude: 'Since this seems to be essentially a matter of terminology, it need not concern us further'.

[18] Matthewson & Demirdache (1995: 69): 'We propose that distinctions between the lexical categories N, V and A … are a universal property of language … We claim that it reflects a deep property of the syntax of Universal Grammar'.

[19] Baker (2003: §1.2) also recognizes the problem with the traditional approach exemplified by (23), and after a discussion of the question of the adjective–verb distinction in Mohawk, he observes: 'The un-answerable question, then, is this: do these differences justify positing a separate category of adjectives in Mohawk after all? Or do we continue to say that Mohawk has only verbs, but concede that there are two subtypes of verbs, intransitive stative verbs and other verbs?' But he does not seem to draw any conclu-sion from this insight.

and pedagogical purposes) be called 'Noun', 'Verb', and 'Adjective', but this labelling convention does not mean that they are concepts of general linguistics.

So, do these terms have a role in general linguistics at all? Or should we use them exclusively as language-particular labels? The answer given by Croft (1991; 2000; 2001) is that the terms *noun*, *verb*, and *adjective* represent 'language universals' or 'typological prototypes':

> Noun, verb and adjective are not categories of particular languages. But noun, verb and adjective are language universals—that is, there are typological prototypes which should be called noun, verb and adjective.

> (Croft 2000: 65)

Croft's important insight is that the widespread sense of universality of nouns, verbs, and adjectives (reflected also in recent generative work like Baker 2003; Chung 2012; Davis et al. 2014) is not wrong, but is not manifested at the level of language-particular classes. It is manifested by universals like those seen in section 2.1, and Croft calls these patterns 'prototypes' and relates them to his 'typological markedness theory' (Croft 2003: ch. 4). But it is odd to describe these correct insights by saying that 'noun' and 'verb' are 'language universals', because a language universal is a statement. And Croft's talk of 'prototypes' has often been misunderstood as claiming that there are no sharp boundaries between the classes of particular languages.[20]

In practical terms, there is a clear answer to the question whether we need nouns, verbs, and adjectives as concepts of general grammar: Yes, we do, because we use these terms as general concepts all the time. We make general statements such as those in (24a–c) about the world's languages or about languages of a particular area. Moreover, even when we talk about particular languages, we often make implicitly comparative statements, as in (25a–c).

(24) a. Tense marking on nouns is much rarer than tense marking on verbs.
　　　b. If a language has dominant SOV order and the genitive follows the governing noun, then the adjective likewise follows the noun (Universal 5 in Greenberg 1963a).
　　　c. In European languages, nouns tend to have obligatory plural marking, whereas plural marking of nouns is often optional in East Asian languages.

(25) a. Japanese has two distinct adjective classes.
　　　b. Verbs do not show person indexing in Modern Swedish.
　　　c. In Yoruba, nouns usually begin with a vowel.

---

[20] For example, Francis & Matthews (2005: 270) say that 'Croft (1991, 2001) defines the prototype for the category "noun" cross-linguistically as a correlation of the semantic class of physical objects with the pragmatic function of reference. Within a given language, some nouns (e.g. those used as modifiers or predicates) may fail to conform to the prototype …' This sounds as if there were more or less prototypical nouns, but this is not what Croft is saying. He says that when physical object words are used in referring function, they are typologically unmarked, and he calls this a 'typological prototype'. (See also Newmeyer 1998: ch. 4 for a similar misunderstanding.)

The statements in (25a–c) are not as precise as they could be if they were made about language-particular categories (Japanese Verby Adjectives, Nouny Adjectives, Swedish Verbs, Yoruba Nouns), but they are highly informative, and we want to continue talking about particular languages in this implicitly comparative way. Thus, in practice, we treat nouns, verbs, and adjectives as cross-linguistic categories. But what do they refer to? Perhaps surprisingly, I suggest that they can be defined as in (26).[21]

(26)   a. A noun is an object-denoting root.
       b. A verb is an action-denoting root.
       d. An adjective is a property-denoting root.

Now one might object that these definitions do not cover everything that we mean by these terms. It may seem that we want to include non-object nouns like *beauty*, non-action verbs like *know*, and maybe non-property adjectives like *royal* as well. In addition, it may seem that we want to include non-root nouns like *play-er*, non-root verbs like *en-large*, and non-root adjectives like *help-ful*.

But while English *know* is a Stative Verb by some English-specific criteria, it is not clear how one could provide a generally applicable (and thus presumably conceptual/semantic) definition that would include it in a general 'verb' category, while not at the same time including many English Adjectives (such as *aware* or *happy*). Thus, it seems that we need to focus on the shared core that all form classes that we generally call 'verbs' have in common: the action meaning. The definitions in (26) give the right results for the universal tendencies in (24) and the language-particular statements in (25), although they are not as informative as one might hope. But since languages differ in many ways, comparative statements are necessarily restricted in scope, and the shared-core definitions in (26a–c) remind us of this.

Now one may grant that the traditional word-class terms are best defined semantically, but one may still ask: couldn't we replace 'root' in (26) by 'lexeme'? Aren't complex words like *play-er*, *en-large*, and *help-ful* object-denoting nouns, action-denoting verbs, and property-denoting adjectives in the same way as roots? Intuitively, this is certainly the case, but there is no clear general understanding of the term 'lexeme', as far as I know. Lexemes are often thought of as abstractions over all forms that only differ in inflectional properties, but since there is no general way of distinguishing between inflection and derivation (e.g. Plank 1994; Spencer 2013),[22] this definition cannot be applied to languages in general. So, it is best to define the traditional word-class terms in terms of roots,

---

[21] A reviewer objected that this terminology will never be adopted in the discipline, and of course it goes against the basic point that word classes are defined grammatically, not semantically (section 2.3–2.4; see n. 12). However, this basic point is valid only for particular languages, and the examples in (24)–(25) are meant to show that in general comparative contexts, linguists actually do use the word-class terms in a notional way.

[22] It is sometimes thought that the difference between inflection and derivation is 'fuzzy', and that this is no different to other distinctions in linguistics (which are all somewhat fuzzy). But there is a crucial difference between a lack of a clear definition and a clear definition with fuzziness. If there is no clear definition, then all we can rely on is stereotypes. I suspect that the inflection–derivation contrast is merely the reflection of traditional ways of describing European languages (in grammars and dictionaries), and has no basis in the structures of languages. For this reason, I do not want to rely on this distinction for the definition of key concepts.

i.e. minimal segmental forms (see Haspelmath 2020 on morphs and their various subtypes, such as affixes and roots).

The definitions in (26) entail that the terms *noun*, *verb*, and *adjective* could alternatively be used in the statement of the universals in section 2.1 and of the macro-types in section 2.2. Thus, we have answered the question raised in these earlier sections about the relevance of these general statements to word classes. They may not be directly relevant to language-particular word classes, but they are highly relevant to the general concepts of noun, verb, and adjective in the sense in which these terms are normally used.

Thus, even though noun, verb, and adjective are (almost by definition)[23] universal semantic types of roots, different languages use different ways of treating such roots in predicative, modifying, and referring functions (as we saw in section 2.2). This is the core of what is generally meant by cross-linguistic word-class variability: languages use function indicators in different ways. But in the next two sections, we will see two further ways in which word classes can be variable: heterosemous root sets (section 2.6) and variation in substantive markers (section 2.7).

2.6 HETEROSEMOUS ROOT SETS

Many languages have pairs such as those listed in (27)–(30), which I call 'heterosemous root sets' here (following Lichtenberk 1991, who introduced the term *heterosemy*, for different meanings of a single element that are associated with different word classes).

(27)   English
    a hammer   *to hammer*
    a mother   *to mother*
    a head   *to head*
    a cook   *to cook*
    a ship   *to ship*
    a dance   *to dance*
    a walk   *to walk*

(28)   Yupik (Mithun 2017b: 163–164)
    amirlu   'cloud'   'be cloudy'
    taqmak/g   'dress'   'put on a dress'
    kuvya   'fishnet'   'fish by driftnetting'

---

[23] Not *quite* by definition, because it is logically possible that a language might lack verbs (= action roots) in the sense of only having object-denoting roots, resorting to complex forms to express actions. For example, 'go' would be expressed as 'use legs', 'eat' as 'use mouth', 'speak' as 'use tongue', 'kill' as 'turn into corpse'. Or, vice versa, everything would ultimately be expressed by event-denoting roots, so 'bird' would be expressed as 'flyer', 'dog' as 'barker', 'woman' as 'birthgiver', 'small' as 'having been reduced', and so on. Thus, the fact that all languages have many object-denoting roots, many action-denoting roots, and many property-denoting roots is an interesting language universal which has not received much attention so far (perhaps because the alternative logical possibility seems so far-fetched).

(29)   Lao (Enfield 2006: 4)
  kèeng3 'soup'   'to make soup of'
  khua5 'fry'   'to fry'
  tôm4 'dish made by boiling' 'to boil'

(30)   Hiw (Oceanic) (François 2017: 333)
  pyë   'a bait'   'to attach a bait'
  veř̄oye 'a fight'   'to fight'
  togekëse 'a game'   'to play'
  vegevage 'speech, language' 'to speak'

For such cases, one might suggest that a single root can occur 'as different word classes'. For example, one might want to say that English *hammer* is a 'category-neutral' root that may 'become a verb' and then have an action meaning, but it might also 'become a noun' and then have an object meaning. This view is often called the 'precategorial' view of such elements (e.g. Evans & Osada 2005: §2.2).

However, it has often been observed that the semantic relationships in such pairs tend to be quite unpredictable and idiosyncratic. And since many languages have corresponding derivational markers (e.g. Spanish *martillo* 'hammer', *martill-ear* 'to hammer'; German *Mutter* 'mother', *be-muttern* 'to mother'), it has been more common to regard such pairs as created by a derivational operation called ZERO-DERIVATION or CONVERSION (e.g. Bauer & Valera Hernández 2005). Unpredictable and idiosyncratic meaning relationships are also otherwise characteristic of derivational relationships.

The problem with the 'derivation/conversion' view is that it presupposes a particular directionality, which is often not evident at all, and there are no good criteria for establishing the direction of derivation: is *a walk* derived from *to walk*? Is *a dance* derived from *to dance*? Is a *a tango* derived from *to tango*? Such questions arise for all languages. (See also Gil 2013a: §4.5.2 for Riau Indonesian in comparison with English.)

Thus, in order to avoid unmotivated decisions, at least in a comparative context, it is best to treat these pairs simply as what they uncontroversially are: heterosemous root sets, i.e. sets of roots that are related by sharing the same shape and having related meanings.[24] Note that we cannot say that elements like *hammer* are 'heterosemous roots', because the approach taken in this chapter requires that by definition, a root (in the comparative sense) can only be associated with one semantic type (an object, an action, or a property). Thus, English *hammer* is a root set, or it could perhaps be said to be a 'super-root', as long as one remembers that a super-root is not a single root but a set of roots.[25]

---

[24] In some languages, there are heterosemous root sets which have to combine with 'stem markers' that clearly indicate the semantic class, e.g. Latin *tim-e(-o)* '(I) fear', *tim-or* '(the) fear' (Lehmann 2008; see also section 6.2 in this volume). For such cases, one might prefer a 'precategorial root' view to a heterosemous root set view, because the root does not occur without a stem marker.

[25] If the term *multicategoriality* is used (François 2017: 299), it must be understood that what is 'multicategorial' is the super-root, not a root in the comparative sense. (Vapnarsky & Veneziano 2017 use *polycategorial* in the same sense.)

But from a language-particular perspective, a rather different treatment of such heterosemous root sets is of course possible: one might say for English that elements like *hammer* and others in (27) are neither Nouns nor Verbs, but represent a third class, perhaps called 'Nomiverbs'. Thus, English might be said to have Nouns like *city* (which can occur after the definite article *the*), Verbs like *steal* (which can occur before the Present-tense 3rd-Singular suffix *-s*), and Nomiverbs like *hammer* (which can occur in both these contexts). In fact, one might say that 'Nomiverb' is the superordinate class, of which Nouns and Verbs are more restricted subclasses which do not occur in all the contexts where unrestricted Nomiverbs occur. This would be a weird way of looking at English heterosemous root sets, but why is it weird? It seems that the main reason is that in a comparative perspective, a concept like 'nomiverb' makes no sense, because comparison always involves nouns, verbs, and adjectives as defined in (26). And even though we want to be faithful to the language-particular peculiarities, in the end we usually assimilate our ways of describing languages to general patterns after all.

In fact, the existence of a Nomiverb class in English has been argued for by Farrell (2001), who claims that words like those in (27) 'are not inherently associated with the syntactic categories "noun" or "verb". Rather, they are associated with meanings that can manifest as either of these categories, by virtue of being compatible with different contextually imposed profiling scenarios' (2001: 128). Clearly, Farrell is more optimistic about the semantic regularity of the relationships than most other linguists. But whatever the merits of such analyses in terms of a Janus-like third class (or even in terms of a superordinate class of which the others are subclasses), they can hardly be the basis of cross-linguistic comparison. For cross-linguistic comparison, we need semantically defined notions, or notions defined in terms of universal construction types (predication, reference, modification).

2.7 MARKERS OF SEMANTIC SUBSTANCE DIFFERENTIATING BETWEEN THE ROOT CLASSES

Many experienced readers will have wondered why my discussion of grammatical marking of nouns, verbs, and adjectives has been limited to function indicators so far (i.e. termifiers, attributivizers, and copulas; recall Table 2.1). To be sure, differences between these three root types have often been motivated by the kinds of behaviours in predicative, modifying, and referring function that we saw above, but of course the literature is much richer.

For example, Dixon (2004: 16) notes that in Cherokee, verbs take tense suffixes but adjectives do not; in Kamaiurá, adjectives have clitic person indexes, while verbs have person prefixes (in some of the moods); and in Korean, adjectives and verbs have different markers for the indicative mood. Similarly, Chung (2012) reports that in Chamorro, verbs but not adjectives allow 'specific external arguments' (see the overview in Haspelmath 2012b: 94). And we can give a straightforward example from English: in predicative position, nouns and verbs take the same copula, but they do not behave in a completely identical way; we can say *Our dog is a labrador* (using the indefinite article on the noun), but not **Our dog is a big* (see (2a)).

So, in addition to function indicators (copulas, termifiers, attributivizers), words may also fall into different classes because they have different substantive markers, i.e. markers such as those in (31).

(31)   substantive markers
        on nouns: articles, plural markers, case markers,
                        possessive indexes, etc.
        on adjectives: comparative markers, degree adverbs, etc.
        on verbs: tense–aspect–mood markers, argument indexes,
                        voice markers, etc.

In many circumstances, these kinds of markers are much more salient for distinguishing nouns, verbs, and adjectives than the function indicators that I focused on in sections 2.1–2.2. For language-particular classification, they are often more immediately relevant than the function indicators. In many languages, nouns are typically accompanied by articles (while verbs and adjectives never are), verbs are obligatorily marked for tense (while nouns never are), and so on.

For example, the English Noun class (excluding the Proper Noun class) can be defined quite easily: a Noun in English is a word that can be preceded by the definite article *the*. This is quite typical, and the literature on word classes is full of such statements. Croft (1991) has a special term for the non-function-indicator markers that can be associated with a class: 'behavioural potential'. Following Ye (2021), I call them *substantive markers* (because they contribute semantic substance, and do not merely point to an atypical propositional act function).

So why did I leave these very important markers aside so far, relegating them to the second-last section? The reason is that they vary too much across languages to be usable for worldwide cross-linguistic comparison. Many languages have articles and plural markers on nouns, but there are also many languages that lack them. Most languages lack comparative marking on adjectives, and verbal marking is very variable, too. Tense marking is quite characteristic of verbs, but there are quite a few languages that only use aspect marking (e.g. Mandarin Chinese), and a few languages do not seem to have tense or aspect marking at all. Thus, we cannot identify word classes across languages in terms of substantive markers, and the same goes for other syntactic properties such as the possibility of 'specific external arguments' in Chamorro.

This issue does not arise for the function indicators, as we saw in section 2.1, and for this reason, they have a privileged role: all languages have predication, modification, and reference, so we can examine how they treat the various semantic root classes in these propositional act functions. As we saw in section 2.2, the classification of words is similar to the classification of argument-marking patterns. Here, too, we have seen a long discussion of universality of syntactic roles like 'subject', and many different criteria have been cited (e.g. occurrence in control constructions, behaviour as reflexive-pronoun antecedents, behaviour in quantifier float constructions). But most of these criteria are not universal, so a general notion of 'subject' must be based on argument marking: The occurrence of flags (ergative or accusative case markers) or of argument indexes, which define A, S, and P, thus allowing us to equate 'subject' with A/S (Haspelmath 2011).

There is thus a solution to the perennial problem of word classes, just as there is a solution to the perennial problem of the subject and other syntactic functions. But the solution

implies a deviation from traditional practice, in that we need to focus our attention on the most basic phenomena: only function indicators in the case of word classes, only argument markers (flags and indexes) in the case of syntactic roles. This is not very interesting from a language-particular perspective, but, as we saw in sections 2.3–2.5, we must keep language-particular analyses separate from general comparative linguistics.

2.8 CONCLUSION

In this chapter, I have drawn some seemingly radical, but unavoidable consequences from an old insight: that different languages have different grammars which are made up of different building blocks. It is obvious to the naked eye that different languages have different words, but once we consider grammatical constructions, one needs a closer look to see that languages also differ in their constructions (maybe less so than in their words, but still quite significantly).

Since we distinguish grammatical classes in a language based on its grammatical constructions, different languages also have different classes, and it is not immediately clear how they can be compared. For example, how should the class of Gendered Tenseless words in Godoberi (section 2.3) be compared with some class of words in Persian, a genderless language, or some class of words in Mandarin, a tenseless language?

The solution to this problem is to base our comparisons not on language-particular constructions or classes (i.e. on language structures), but on comparable forms, with their meanings and shapes (i.e. on conceptual and phonetic substance). As seen in section 2.1, we can formulate universals based on cross-linguistic comparison in terms of object roots (or 'nouns', section 2.5), action roots (or 'verbs'), and property roots (or 'adjectives'). These are defined in terms of universally available concepts, and a root is simply defined as a minimal form (a segment sequence not consisting of other forms, Haspelmath (2020)). In addition, we need the comparative concept of *function indicator* (a cover term for termifiers, attributivizers, and copulas), which is defined as a marker (= a bound form that is not a root) which signals the propositional act function (i.e. predication, modification, reference) of its host. These core comparative concepts are independent of language-particular constructions and allow us to formulate universals (as in section 2.1) and to identify some macro-types in terms of coexpression patterns of function indicators.

There is a long tradition in comparative grammar of identifying some languages as being 'flexible' in their word classes and others as more 'differentiated', but, as we saw in section 2.2, it is quite difficult to make general statements across the three propositional act functions. Clearly, most (or all?) languages are not fully differentiating, but we do not know whether there are implicational links among coexpression patterns of different propositional act functions. Languages that completely lack any function indicators and that are thus fully acategorial do not seem to exist.

Much past work on word classes across languages has not singled out function indicators as I have done in this chapter (following Croft 1991; 2000), but has attempted to take all kinds of properties of different word classes into account. This approach does not work, because it is only function indicators that are comparable across languages (section 2.7). There seem to be some tendencies in some kinds of substantive markers (in particular, tense markers

tend to be restricted to action words and similar words), but these do not lend themselves so easily to cross-linguistic generalizations.

Finally, I should note that not all linguists agree with Sapir and others that different languages have different classes (section 2.3). Instead, these linguists think that by examining languages more closely, we will eventually find that they can be described in terms of a universal set of innate building blocks (e.g. Baker 2003; Chung 2012; Davis et al. 2014; Panagiotidis 2014).[26] However, there is nothing close to a consensus on what this set of innate building blocks might be, so, in practice, this approach is difficult to apply to a larger set of languages. Simply hypothesizing that all languages have the categories noun, verb, and adjective because these are part of an innate grammar blueprint ('universal grammar'), as is done by Chung (2012) and others, will not give reliable results as long as we do not know much about what is innate, and it is prone to yielding results that are coloured by expectations based on European languages (Haspelmath 2012b).

---

[26] Gil's (2000) proposal is similar to these generative authors in positing a universal set of building blocks, but with no reference to the usual categories of noun, verb, adjective. Instead, Gil makes reference to simple combinatorial properties that could be applied to a larger set of languages. This is an interesting alternative that is worth considering.

References

Baker, Mark C. 2003. *Lexical categories: Verbs, nouns, and adjectives*. Cambridge: Cambridge University Press.

Baptista, Marlyse. 2013. Cape Verdean Creole of Brava structure dataset. In Michaelis, Susanne Maria & Maurer, Philippe & Haspelmath, Martin & Huber, Magnus (eds.), *Atlas of Pidgin and Creole Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (https://apics-online.info/contributions/56)

Bauer, Laurie & Valera Hernández, Salvador (eds.). 2005. *Approaches to conversion / zero-derivation*. Münster: Waxmann.

Beck, David. 2002. *The typology of parts of speech systems: The markedness of adjectives*. London: Routledge.

Bickel, Balthasar. 2011. Grammatical relations typology. In Song, Jae Jung (ed.), *The Oxford handbook of linguistic typology*, 399–444. Oxford: Oxford University Press.

Bloomfield, Leonard. 1933. *Language*. New York: H. Holt and Company.

Chumakina, Marina. 2018. Atributiv v arčinskom jazyke [Атрибутив в арчинском языке]. *Rhema [Рема]* 2018(4). (doi:10.31862/2500-2953-2018-4-166-189)

Chung, Sandra. 2012. Are lexical categories universal? The view from Chamorro. *Theoretical Linguistics* 38(1–2). 1–56. (doi:10.1515/tl-2012-0001)

Chung, Sandra. 2012. Are lexical categories universal? The view from Chamorro. *Theoretical Linguistics* 38(1–2). 1–56. (doi:10.1515/tl-2012-0001)

Croft, William & van Lier, Eva. 2012. Language universals without universal categories. *Theoretical Linguistics* 38(1–2). (doi:10.1515/tl-2012-0002)

Croft, William. 1991. *Syntactic categories and grammatical relations: The cognitive organization of information*. Chicago: University of Chicago Press.

Croft, William. 2000. Parts of speech as language universals and as language-particular categories. In Vogel, Petra M. & Comrie, Bernard (eds.), *Approaches to the typology of word classes*, 65–102. Berlin: Mouton de Gruyter.

Croft, William. 2001. *Radical Construction Grammar: Syntactic theory in typological perspective*. Oxford: Oxford University Press.

Croft, William. 2009. Methods for finding universals in syntax. In Scalise, Sergio & Magni, Elisabetta & Bisetto, Antonietta (eds.), *Universals of language today*, 145–164. Dordrecht: Springer.

Croft, William. 2022. *Morphosyntax: Constructions of the world's languages*. Cambridge: Cambridge University Press

Croft, William. 2023. Word classes in Radical Construction Grammar. In: Eva van Lier (ed.) *The Oxford handbook of word classes*.

Davis, Henry & Gillon, Carrie & Matthewson, Lisa. 2014. How to investigate linguistic diversity: Lessons from the Pacific Northwest. *Language* 90(4). e180–e226.

Dik, Simon C. 1997. *The theory of Functional Grammar: 2 volumes*. Berlin: Mouton de Gruyter.

Dixon, R. M. W. 2010. *Basic linguistic theory: Volume 1*. Oxford: Oxford University Press.

Dryer, Matthew S. 2007. Clause types. In Shopen, Timothy (ed.), *Language typology and syntactic description, Vol. 1*, 224–275. Second Edition. Cambridge: Cambridge University Press.

Enfield, N. J. 2006. Heterosemy and the grammar-lexicon trade-off. In Ameka, Felix K. & Dench, Alan & Evans, Nicholas (eds.), *Catching language: The standing challenge of grammar writing*, 297–320. Berlin: Mouton de Gruyter.)

Evans, Nicholas & Osada, Toshiki. 2005. Mundari: The myth of a language without word classes. *Linguistic Typology* 9(3). 351–390. (doi:10.1515/lity.2005.9.3.351)

Evans, Nicholas. 2000. Word classes in the world's languages. In Booij, Geert & Lehmann, Christian & Mugdan, Joachim (eds.), *Morphology: A handbook on inflection and word Formation, vol. 1* (Handbücher zur Sprach- und Kommunikationswissenschaft), 708–723. Berlin: Mouton de Gruyter.

Farrell, Patrick. 2001. Functional shift as category underspecification. *English Language & Linguistics* 5(1). 109–130. (doi:10.1017/S1360674301000156)

Fassberg, Steven E. 2019. Modern Western Aramaic. In Huehnergard, John & Pat-El, Na'ama (eds.), *The Semitic languages*, 632–652. London: Routledge.

Foley, William A. 2023. Variation in word class systems and their interaction with other grammatical properties. In van Lier, Eva (ed.), *Oxford handbook of word classes.* Oxford: Oxford University Press.

Francis, Elaine J. & Matthews, Stephen. 2005. A multi-dimensional approach to the category 'verb' in Cantonese. *Journal of Linguistics*. Cambridge University Press 41(2). 269–305. (doi:10.1017/S0022226705003270)

François, Alexandre. 2017. The economy of word classes in Hiw, Vanuatu: Grammatically flexible, lexically rigid. *Studies in Language* 41(2). 294–357. (doi:10.1075/sl.41.2.03fra)

Garvin, Paul L. 1951. Kutenai IV: Word classes. *International Journal of American Linguistics* 17(2). 84–97. (doi:10.1086/464111)

Gerdts, Donna B. 2023. Word classes in Salishan languages. In van Lier, Eva (ed.), *Oxford handbook of word classes.* Oxford: Oxford University Press.

Gibson, Edward & Futrell, Richard & Piandadosi, Steven T. & Dautriche, Isabelle & Mahowald, Kyle & Bergen, Leon & Levy, Roger. 2019. How efficiency shapes human language. *Trends in Cognitive Sciences* 23(5). 389–407. (doi:10.1016/j.tics.2019.02.003)

Gil, David. 1993. Syntactic categories in Tagalog. *Pan-Asiatic Linguistics: Proceedings of the Third International Symposium on Languages and Linguistics Bangkok January 8-10 1991, ed. Sudaporn Luksaneeyanawin*, 1136–1150. Bangkok: Chulalongkorn University Press.

Gil, David. 2000. Syntactic categories, cross-Linguistic variation and Universal Grammar. In Vogel, Petra M. & Comrie, Bernard (eds.), *Approaches to the typology of word classes*, 173–216. Berlin: Mouton de Gruyter.

Gil, David. 2005. Adjectives without nouns. In Haspelmath, Martin & Dryer, Matthew S & Gil, David & Comrie, Bernard (eds.), *The world atlas of language structures*. Oxford: Oxford University Press. (https://wals.info/chapter/61)

Gil, David. 2013. Riau Indonesian: A language without nouns and verbs. In Rijkhoff, Jan & van Lier, Eva (eds.), *Flexible word classes: Typological studies of underspecified parts of speech*, 89–130. Oxford: Oxford University Press.

Gordon, Lynn. 1986. *Maricopa morphology and syntax*. Berkeley: University of California Press.

Haspelmath, Martin. 1993. *A grammar of Lezgian* (Mouton Grammar Library 9). Berlin: Mouton de Gruyter.

Haspelmath, Martin. 2005. Argument marking in ditransitive alignment types. *Linguistic Discovery* 3(1). 1–21.

Haspelmath, Martin. 2012a. How to compare major word-classes across the world's languages. In Graf, Thomas & Paperno, Denis & Szabolcsi, Anna & Tellings, Jos (eds.), *Theories of everything: in honor of Edward Keenan* (UCLA Working Papers in Linguistics 17), 109–130. Los Angeles: UCLA.

Haspelmath, Martin. 2012b. Escaping ethnocentrism in the study of word-class universals. *Theoretical Linguistics* 38(1–2). (doi:10.1515/tl-2012-0004)

Haspelmath, Martin. 2014. Descriptive hypothesis testing is distinct from comparative hypothesis testing: Commentary on Davis, Gillon, and Matthewson. *Language* 90(4). e250–e257. (doi:10.1353/lan.2014.0071)

Haspelmath, Martin. 2015. Ditransitive constructions. *Annual Review of Linguistics* 1. 19–41. (doi:10.1146/annurev-linguist-030514-125204)

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. Explaining grammatical coding asymmetries: Form-frequency correspondences and predictability. *Journal of Linguistics* 57(3). 605-633. (doi:10.1017/S0022226720000535)

Himmelmann, Nikolaus. 2008. Lexical categories and voice in Tagalog. In Austin, Peter & Musgrave, Simon (eds.), *Voice and grammatical relations in Austronesian languages*. Stanford: CSLI Publications. (https://research.monash.edu/en/publications/voice-and-grammatical-relations-in-austronesian-languages)

Kibrik, Alexandr (ed.). 1996. *Godoberi*. München: Lincom Europa.

Launey, Michel. 1994. *Une grammaire omniprédicative: Essai sur la morphosyntaxe du nahuatl classique*. Paris: CNRS Editions.

Lehmann, Christian. 2008. Roots, stems and word classes. *Studies in Language* 32(3). 546–567. (doi:10.1075/sl.32.3.04leh)

Lichtenberk, Frantisek. 1991. Semantic change and heterosemy in grammaticalization. *Language*. Linguistic Society of America 67(3). 475–509. (doi:10.2307/415035)

Lyons, John. 1968. *Introduction to theoretical linguistics*. Cambridge: Cambridge University Press.

Lyons, John. 1977. *Semantics*. Cambridge: Cambridge University Press.

Malchukov, Andrej L. 2023. Transcategorial operations: A Generalized Scale Model approach. In van Lier, Eva (ed.), *Oxford handbook of word classes*. Oxford: Oxford University Press.

Matthewson, Lisa & Demirdache, Hamida. 1995. Syntactic categories in St'át'imcets (Lillooet Salish). *University of British Columbia Working Papers in Linguistics:30th International Conference on Salish and Neighbouring Languages)* 3. 69–75.

Michaelis, Susanne Maria & Rosalie, Marcel. 2013. Seychelles Creole structure dataset. In Michaelis, Susanne Maria & Maurer, Philippe & Haspelmath, Martin & Huber, Magnus (eds.), *Atlas of Pidgin and Creole Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (https://apics-online.info/contributions/56)

Mithun, Marianne. 2017. Polycategoriality and zero derivation: Insights from Central Alaskan Yup'ik Eskimo. In Vapnarsky, Valentina & Veneziano, Edy (eds.), *Lexical polycategoriality: Cross-linguistic, cross-theoretical and language acquisition approaches*, 155–174. Amsterdam: Benjamins.

Mosel, Ulrike. 2023. Word classes in Austronesian languages. In van Lier, Eva (ed.), *Oxford handbook of word classes*. Oxford: Oxford University Press.

Newmeyer, Frederick J. 1998. *Language form and language function*. Cambridge, MA: MIT Press.

Nordhoff, Sebastian. 2013. Jack of all trades: The Sri Lanka Malay flexible adjective. In Rijkhoff, Jan & van Lier, Eva (eds.), *Flexible word classes: typological studies of underspecified parts of speech*, 247–274. Oxford: Oxford University Press.

Panagiotidis, Phoevos. 2014. *Categorial features: A generative theory of word class categories*. Cambridge: Cambridge University Press.

Plank, Frans. 1994. Inflection and derivation. In Asher, R.E. (ed.), *Encyclopedia of Language and Linguistics*, 1671–1678. Oxford: Pergamon.

Ramat, Paolo. 2023. Word classes in the history of linguistics. In van Lier, Eva (ed.), *Oxford handbook of word classes*. Oxford: Oxford University Press.

Rijkhoff, Jan & van Lier, Eva (eds.). 2013. *Flexible word classes: Typological studies of underspecified parts of speech*. Oxford: Oxford University Press.

Robins, R. H. 1952. Noun and verb in universal grammar. *Language* 28(3). 289–298. (doi:10.2307/410101)

Sapir, Edward. 1921. *Language: An introduction to the study of speech*. New York: Harcourt, Brace & Co.

Sasse, Hans-Jürgen. 1993. Syntactic categories and subcategories. In Jacobs, Joachim & von Stechow, Arnim & Sternefeld, Wolfgang & Vennemann, Theo (eds.), *Syntax: An international handbook of contemporary research*, vol. 2, 646–686. Berlin: De Gruyter.

Sasse, Hans-Jürgen. 2015. Syntactic categories and subcategories. In Kiss, Tibor & Alexiadou, Artemis (eds.), *Syntax: Theory and analysis* (Handbücher Zur Sprach- Und Kommunikationswissenschaft (HSK) 42), vol. 1, 158–217. Berlin, Boston: De Gruyter.

Schachter, Paul & Shopen, Timothy. 2007. Parts-of-speech systems. In Shopen, Timothy (ed.), *Language typology and syntactic description. Vol. I: Clause structure*, 1–60. Cambridge: Cambridge University Press.

Siewierska, Anna & Bakker, Dik. 2009. Case and alternative strategies: Word order and agreement marking. In Malchukov, Andrej & Spencer, Andrew (eds.), *The Oxford handbook of case*, 290–303. Oxford: Oxford University Press.

Spencer, Andrew. 2013. *Lexical relatedness*. Oxford: Oxford University Press.

Stassen, Leon. 1997. *Intransitive predication*. Oxford: Oxford University Press.

Vapnarsky, Valentina & Veneziano, Edy (eds.). 2017. *Lexical polycategoriality: Cross-linguistic, cross-theoretical and language acquisition approaches*. Amsterdam: Benjamins.

Viljoen, Melanie Helen. 2013. *A grammatical description of the Buwal language*. Melbourne: LaTrobe University. (PhD dissertation.)

Vogel, Petra M. & Comrie, Bernard (eds.). 2000. *Approaches to the typology of word classes*. Berlin: Mouton de Gruyter.

Yap, Foong Ha & Grunow-Hårsta, Karen & Wrona, Janick. 2011. Introduction: Nominalization strategies in Asian languages. In Yap, Foong Ha & Grunow-Hårsta, Karen & Wrona, Janick (eds.), *Nominalization in Asian Languages: Diachronic and typological perspectives*, 1–60. Amsterdam: Benjamins.

Ye, Jingting. 2021. *Property words and adjective subclasses in the world's languages*. Leipzig: Leipzig University. (Ph D dissertation.)

Zipf, George Kingsley. 1935. *The psycho-biology of language: An introduction to dynamic philology*. Cambridge: MIT Press.