C17

CHAPTER 17

NOUNS

JAN RIJKHOFF

C17S1

17.1 Introduction

C17P1 The prototypical noun is a basic, underived content word, which serves as the head of a noun phrase that is typically used to refer to one or more concrete, physical objects like cups, cats, or cars ('first order entities'; Lyons 1977: 442). However, nouns are also employed to talk about masses (*water*$_N$), places (*valley*$_N$), collectives (*crowd*$_N$) and emotions (*love*$_N$, *fear*$_N$), as well as events ('second order entities') like *wedding*$_N$ or *meeting*$_N$, propositions or possible facts ('third order entities') like *belief*$_N$ or *opinion*$_N$, and speech acts ('fourth order entities') like *question*$_N$ or *comment*$_N$. For reasons explained below (see in particular section 17.3), the current chapter is mostly concerned with basic, unmarked nouns that are used to refer to a concrete object in the physical world.

C17P2 This chapter offers an overview of members of the word-class Noun from a typologically informed, cross-linguistic perspective (see Part I of this volume for other approaches to word classes). It is sometimes assumed that a distinct lexical category Noun is attested in every natural human language (Sapir 1921: 119; Whaley 1997: 32; Croft 2003: 183; Chung 2012), but this appears not to be the case. It has been argued, for example, that nouns cannot be distinguished syntactically from other major word classes like verbs or adjectives in the Polynesian languages Samoan and Tongan (Mosel & Hovdhaugen 1992: 73; Broschart 1997; see also, e.g. Himmelmann 2005a: 128). Furthermore, it has been noted that reference to a concrete object in Oneida and other indigenous North American languages commonly involves the use of verbal forms or constructions (Michelson 1990: 76; Mithun 1999: 60–61, 82; Abbott 2000: 48).

C17P3 The remainder of this chapter proceeds as follows. Section 17.2 is concerned with the status of the word-class Noun as a cross-linguistic lexical category. Section 17.3 presents a cross-linguistic classification of the nominal lexemes that are central in this chapter: nouns that are used to talk about a spatial object in the external world, also known as 'concrete nouns' or 'first order nouns'. Section 17.4 is concerned with lexemes that do not fit easily in the classification of nominal subcategories presented in section 17.3. Section 17.5 offers a brief overview of certain (other) semantic, morphological, phonological, or cognitive properties of nouns. The chapter ends with a conclusion (section 17.6).

C17S2

17.2 Nouns as a cross-linguistic lexical category

C17P4 It is important to note at the outset that the problem of identifying members of the cross-linguistic word-class Noun is not so much trying to identify words in the various languages whose grammatical properties are completely identical (no two linguistic forms or constructions are exactly the same grammatically, not even in a single language; Gross 1979), but rather making sure that the units are similar enough to allow for a responsible cross-linguistic comparison (Rijkhoff 2016: 333).

C17P5 In the European linguistic tradition, nouns are often defined in terms of certain formal, morphosyntactic features, but form-based characterizations of members of linguistic categories are too language-specific to be useful for an overall, cross-linguistic comparison. For example, one of the most conspicuous properties of (count) nouns in European languages is the fact that they can be inflected for plural number (*cat*$_{N}$ vs *cat-s*$_{N}$). From a cross-linguistic perspective, however, this is not a suitable criterion, because number marking is absent or at best optional (but see section 17.3.2.3) in many, possibly even most of the world's languages (Rijkhoff 2004: 110–111, 146–152). Besides, even languages with compulsory number marking have a significant number of nouns that do not inflect for plural number, such as mass nouns (*gold-s) and abstract nouns (*courage-s).

C17P6 Semantic criteria have also turned out to be unhelpful (Rijkhoff 2007: 711). As mentioned, prototypical nouns are associated with concrete, spatial objects, but many nouns are used to talk about non-spatial entities such as events (*meeting*$_{N}$), feelings (*love*$_{N}$), and other abstract entities (*linguistics*$_{N}$). Furthermore, a pre-linguistic notion that is lexicalized as a noun in one language may be expressed in another language as a verb (Evans 2000; Urban 2012: 152–168) or an adjective (Wierzbicka 1986: 354–355; Rijkhoff 2015: 645).[1] As noted by Lehmann (1990: 181), 'the choice of one lexical categorization instead of another has far-reaching consequences for the whole linguistic system' (see also Hengeveld 2013; Chapter 6 in this volume).

C17P7 In sum, whereas word classes that are defined in terms of morphosyntactic properties (such as the ability to be inflected for number, definiteness, or case) are too narrow in that they do not cover all the relevant words in languages across the globe, semantically defined lexical categories are typically too wide, because they include words that belong to different parts of speech in the various languages (Rijkhoff 2009, 2016).

C17P8 In this chapter, I will use Hengeveld's (1992b, 2013) approach to word classes as the point of departure, one of the main reasons being that his classification of parts-of-speech systems is based on descriptions of word classes in a representative sample of 50 languages (see also

[1] The fact that across languages the same notion may be lexicalized as a member of a different word class has been discussed in various publications, notably Lehmann (2010: 166–171; see also, e.g. Croft 1993). See, for example, Ross (1998a: 90) and Devos (2008: 136) on the way adjectival notions like English *pretty*$_{A}$ or *silent*$_{A}$ are expressed as relative clauses headed by a stative verb like *tikiraoi*$_{V}$ 'be_pretty' in Kiribati (an Oceanic language) or as genitival noun phrases headed by an abstract noun like *-búúli*$_{N}$ 'silence' in Makwe (a Bantu language), respectively (see also, e.g. Dixon 1982).

Chapter 10 in this volume).[2] Notice furthermore that Hengeveld's classification is included in Bisang's (2015: 553) comparison of the two most comprehensive typological approaches to word-class categorization (the other approach was proposed in Croft 2000; see also Chapter 11 in this volume).

C17P9 The observation that nouns and other major lexical word classes appear to be language-particular categories regarding their formal or semantic properties does not mean they cannot be matched across languages. Members of the cross-linguistic word-class Noun may be compared on functional and structural grounds in terms of likeness (Rijkhoff 2016; McGregor 2019a: 208), as in Hengeveld's (here: slightly reformulated) definition of nouns as part of his classification of parts-of-speech systems (Hengeveld 1992b: 68): A DEDICATED NOUN IS A CONTENT WORD THAT CAN BE IMMEDIATELY USED AS THE HEAD OF A NOUN PHRASE.[3] Thus, English words like *think*$_{V}$ (a verb), *tall*$_{A}$ (an adjective) or *today*$_{Adv}$ (an adverb) can only serve as the head of a noun phrase when they appear as derived nouns (such as *think-er*$_{N}$ or *tall-ness*$_{N}$).

C17P10 This definition leaves open two possibilities, both of which are attested in natural human languages. First, it allows for the possibility that there are languages without a separate, dedicated class of nouns (see section 17.1): (a) languages with a flexible word class like Samoan (see Chapter 3 in this volume), where—syntactically speaking—a separate category of nouns is absent, and (b) languages such as Oneida, whose speakers generally use verbs or verb-based expressions to talk about concrete, physical objects. Second, the definition also permits nouns not to appear as the head of a noun phrase, but by themselves when they serve as a non-verbal clausal predicate (see n. 3), as in the case of Dutch professional names like *leraar*$_{N}$ 'teacher' or *loodgieter*$_{N}$ 'plumber', regardless of the number of the subject noun phrase (NP).

C17P11 Dutch (Germanic): the non-verbal predicate *loodgieter*$_{N}$ is just a noun (any form of inflection or modification would turn the bare noun into the head of a noun phrase):

(1) *Ben en Peter zijn loodgieter*$_{N}$
Ben and Peter be:PL.PRES plumber
'Ben and Peter are plumbers'.

C17P12 It is also possible to have a full noun phrase instead of just a bare noun as the non-verbal predicate here, as in (2).

C17P13 Dutch: the inflected noun *loodgieters*$_{N}$ is the head of a noun phrase:

(2) *Ben en Peter zijn* [(twee ervaren) *loodgieter-s*]$_{NP}$
Ben and Peter be:PL.PRES (two experienced) plumber-PL
'Ben and Peter are (two experienced) plumbers'.

C17P14 The difference between having just a nominal lexeme (N) or a phrasal construction (NP) as the non-verbal clausal predicate is reflected in a subtle semantic difference, where the

[2] See Majid & Levinson (2010) and Dahl (2015) on the influence of languages spoken in Western, Educated, Industrialized, Rich, and Democratic (WEIRD) societies in linguistic studies.

[3] Non-verbal lexemes like nouns, adjectives and adverbs may also be used in predicative function, i.e. as the head of the clause, as in *Marianne is professor*$_{N}$/*happy*$_{A}$/*here*$_{Adv}$ (see Hengeveld 1992 for further details).

predicate noun seems to appear as a semantically stripped version of the lexicalized meaning of the noun, affecting the way the nominal property is represented in the spatial dimension (see section 17.3.1 on the 'mode of being' or *Seinsart* of a noun). The difference has been described in terms of individuation in that the bare (number-neutral or transnumeral) predicate noun is said to 'not to individuate' (Geerts et al. 1984: 145). The same seems to hold true for incorporated nouns, such as 'book' in *book-shop*$_{N}$ (Rijkhoff 2004: 56–57; on this topic, see also Wackernagel 1920: 84; Hundt 2016; Görgülü 2018; see Mithun 1984 for a discussion of various types of incorporation).[4]

C17S3

17.3 A Cross-linguistic classification of nouns

C17P15 It was mentioned in section 17.2 that the objects in a cross-linguistic comparison must be sufficiently similar to warrant a responsible investigation. So as to be able to do a more focused comparison between members of the cross-linguistic category Noun, another criterion has been added to the general definition of nouns provided earlier, ensuring even better that nouns in different languages are studied against each other in terms of the same properties (Song 2001: 11). On the assumption that there are concrete objects in the physical world ('common sense metaphysics'), these objects are used as external reference points (TERTIA COMPARATIONIS) to make certain that the nouns in our cross-linguistic comparison are sufficiently similar in function, meaning and form (Rijkhoff 2008: 732; cf. Ursini & Acquaviva 2019). In other words, even though nouns can be used to talk about many different kinds of entities (section 17.1), ranging from concrete physical objects (*cup*, *car*) to very abstract entities (*objection*, *trapezium*, *courage*) and anything in between (*game*, *wedding*), the current chapter will mostly focus on nouns with the widest cross-linguistic distribution: underived nouns that are used to refer to concrete objects in the external, physical world. The fact that abstract nouns are rare or generally absent in certain languages is mentioned in, for example, Dixon (1980: 272) and Corbett (2000: 87).

C17P16 Due to this restriction, nouns denoting abstract and higher-order entities, like *fear*, *game*, or *opinion*, remain undiscussed here, as well as derived and other morphologically complex nominal forms like compounds and nominalizations (Koptjevskaja-Tamm 1993; Malchukov 2006; Bauer 2017). This also includes cases where a language employs a complex noun where another language would use a non-complex noun for the same object ~~on~~ the physical world. A case in point is the Papuan language Kalam, which 'has a fair supply of nominal morphemes', but lacks a simple nominal lexeme for *child* and certain other 'conceptual categories which one might expect to be universal' (Pawley 1993: 99f.; see also e.g. Goddard 2001: 1194; Lucy 2010: 268; Urban 2012).

[4] The bare predicate noun and the incorporated noun discussed in this chapter should not be confused with the category of 'bare nouns' in formal semantics or syntax, where 'bare nouns' also include nouns inflected for e.g. plural number (Delfitto 2006). In the formalist tradition, the term 'bare noun' (or 'bare noun phrase') is commonly used in the context of the Determiner Phrase (DP) analysis of noun phrases and basically refers to a noun or noun phrase without a quantifier or determiner (see de Swart & Zwarts 2009).

C17P17 Furthermore, I will ignore special cases such as light nouns, i.e. nouns which under certain conditions lose some of their referential potential (Simone & Masini 2014), generic nouns (Dixon 1980: 272–273; Vittrant 2005), numerative nouns (Nurmio & Willis 2016) as well as PLURALIA TANTUM like *scissors*$_N$ or *pajamas*$_N$, which have no singular form (Wisniewski 2010: 181–184; Corbett 2019), and SINGULARIA TANTUM such as *jewellery*$_N$ or *furniture*$_N$, which have no plural form.[5] The reason that a noun like *furniture*$_N$ has no singular–plural opposition is that things of different kinds cannot be counted together (Frege 1950: 62; Mithun 1999: 92–93; Wierzbicka 1985: 320–321). Other types of nouns that have been regarded as SINGULARIA TANTUM include verbal nouns/infinitives (Dutch *het verkopen van* … lit. 'the selling of …'), abstract nouns (*thirst*$_N$, *love*$_N$, *linguistics*$_N$, *politics*$_N$), proper names (*Mark*$_N$, *Lisa*$_N$; see Van Langendonck 2007, Pina-Cabral 2015, Hough 2016, Mackenzie 2018), and names for chemical elements (*helium*$_N$).

C17P18 Although mass nouns are normally not used to talk about concrete objects in the external world (and would therefore strictly speaking fall outside the scope of this chapter), they are discussed in this chapter for reasons that will become clear later (sections 17.3.3.3 and 17.3.3.4).

C17P19 Finally, even though this chapter focuses on basic, unmarked nouns that are used to refer to concrete objects in the physical world, it is interesting to note that derived and higher-order nouns have a strong tendency to model themselves grammatically after the underived first order nouns in the same language. For example, the noun *arrival*$_N$ is derived from the verb *arrive*$_V$ and behaves just like a basic count noun (compare *one arrival—many arrivals* and *one umbrella—many umbrellas*), whereas the noun derived from the verb *sleep*$_V$ behaves like a mass noun (*much sleep*$_N$, *much water*$_N$). This is captured by the Principles of Formal and Semantic Adjustment (Dik 1997, Part 2: 20, 158–164); other manifestations of this principle are discussed in Lehmann (1990), who also refers to Benveniste (1957).

AQ: Please see text now as corrected re numbers in headings and amend accordingly – too many big changes for me to be able to verify [20 was used in places instead of 17& numbers were skipped/out of sequence]

OK

C17S4

17.3.1 Classifying nouns: Introducing the semantic features Shape and Homogeneity

C17P20 Concrete objects in the physical world, like cats or umbrellas, are typically characterized by a perceptual spatial boundary (i.e. they have a shape) and consist of connected parts (i.e. they have an internal structure) that stay together when the object is moved (Acquaviva 2008: 99–101; Levin & Rappaport Hovav 2011). By contrast, liquids and other masses lack a definite spatial outline and are homogeneous entities, i.e. they typically consist of non-individuated units or portions (rather than connected parts) that are all of the same kind, e.g. drops of a liquid or slices of a substance. It turns out that the properties SHAPE and HOMOGENEITY (or 'Likepartedness') can be felicitously exploited to characterize nouns that are employed in languages across the globe to talk about spatial entities (Rijkhoff 2004: 28–59; Rijkhoff 2008).

[5] SINGULARIA TANTUM such as *furniture* or *jewellery* are sometimes called 'collective mass nouns', 'object mass nouns', 'fake mass nouns', 'collective aggregates', or simply 'collectives', but they should not be confused with the collective nouns like English 'crew' or 'swarm' (which can all be pluralized) or the genuine mass nouns discussed in sections 17.3.2.2 and 17.3.3.3.

C17P21 It is important to emphasize that the features SHAPE and HOMOGENEITY are here used for a linguistic classification of nouns rather than for an ontological classification of physical objects, so there need not be a direct relationship between noun type and (real world) entity type. In other words, we use ontological features of spatial entities for a linguistic classification of nouns, which may or may not have lexicalized these features. This is, of course, precisely the reason that different kinds of nouns in languages across the globe can be used to talk about the same physical object (or *Sein*-correlate) in the external world (Rijkhoff 2004: 55–56). As noted by Levin & Rappaport Hovav (2011: 1): 'since any entity in the world is constituted of many attributes of which a noun lexicalizes only some, two nouns may refer to the same entity, but in lexicalizing different attributes, they may construe it as an entity in different ways'. The fact that the same real-world object (such as an umbrella) can be referred to by using different kinds of nouns may even reflect 'differences in the ontological beliefs the speakers of these languages hold about the referents of nouns' (Foley 1997: 231; more on the topic of linguistic relativity in section 17.3.3). The classification of nominal subcategories in Figure 17.1 shows that the six nominal subtypes have different values for the features Shape and Homogeneity: positive (+), negative (–) or neutral (no value).

C17P22 As only entities with a definite outline can be counted directly, one may assume that an English noun like *umbrella*$_N$, which can be in a direct construction with a numeral ('two umbrellas'), has a positive value for the feature Shape (+Shape). In this case, a positive value for the semantic feature Shape matches a conspicuous ontological property of an umbrella in the real world.

C17P23 The same English word (*umbrella*$_N$) is characterized by a negative value for the semantic feature Homogeneity. Liquids like water and other homogeneous entities do not come in distinct, individuated units but consist of portions of the same kind. In English, these portions can be referred to by the same name as the whole. For example, when we pour some water from a pitcher in a glass, both the liquid in the glass and the remaining liquid in the pitcher are still referred to as 'water'. But since we do not speak of two umbrellas when we break one in two pieces, we may infer that the English noun *umbrell*a is characterized by a negative value for the semantic feature Homogeneity (–Homogeneity; notice that children up to age 4 count parts as wholes; Shipley & Shepperson 1990), in this case, too, the value of the semantic feature corresponds to an ontological property of umbrellas in the real world.

C17P24 If we now compare the English noun *umbrella* (which is classified as a SINGULAR OBJECT NOUN below) with its translational equivalent in Thai *rôm*$_N$ 'umbrella(s)', we find that Thai *rôm*$_N$ is a number-neutral or transnumeral noun that cannot enter into a direct construction with a cardinal numeral.[6] This is because Thai nouns, like other languages with true numeral (or 'sortal') classifiers (see n. 13), are deemed to 'purely denote concepts and, for this reason, are incompatible with direct quantification' (Hundius & Kölver 1983: 166), which suggests that the Thai noun *rôm*$_N$ 'umbrella(s)' is characterized by a negative value for the feature Shape (–Shape). This does not mean that speakers of Thai do not know that an umbrella in the physical world is a discrete object, but rather that this particular piece of encyclopaedic knowledge is not part of the lexicalized meaning of *rôm*$_N$ 'umbrella(s)' (Rijkhoff 2008: 747; for a similar point see e.g. Unterbeck 1993).

[6] Notice that the term 'general number' is sometimes used in connection with different kinds of transnumeral nouns (including, e.g. set nouns and sort nouns, discussed in sections 17.2.3.2 and 17.3.3.1), when 'the meaning of the noun can be expressed without reference to number' (Corbett 2000: 10).

C17P25 Before *rôm*$_{N}$ and other first order Thai nouns like *mánaaw*$_{N}$ 'lemon(s)' can be counted, they must be provided with a numeral or sortal classifier (CLF), which serves to individualize the property denoted by such nouns (Lyons 1977: 462), as in:

Thai (Austro-Tai) (Hundius & Kölver 1983: 172, 167)

(3) *rôm* *săam* *khan*
umbrella(s) three CLF:LONG, HANDLED OBJECT
'three umbrellas'

(4) *mánaaw* *săam* *lûuk*
lemon(s) three CLF:FRUIT
'three lemons'

C17P26 The sortal classifier is required, because the lexicalized meaning of nouns like *rôm*$_{N}$ and *mánaaw*$_{N}$, which are categorized as SORT NOUNS in Figure 17.1, does not seem to include the notion of spatial boundedness or discreteness (Hundius & Kölver 1983; Bisang 1999; Vittrant 2005: 135). The employment of sortal classifiers can allow speakers to emphasize different aspects of an entity (Adams 1989: 3). This is shown in (5) from Mandarin Chinese, where the classifier *duo*$_{CLF}$ indicates that the noun phrase is used to refer to the bud of a plant or flower, while *zhu*$_{CLF}$ makes clear that the noun phrase with the same head noun (*hua*$_{N}$) is used to refer to the plant itself. In either case, the classifier coerces the noun to have a certain reading, depending on the information 'entailed' in the classifier (on coercion, see also e.g. Audring & Booij 2016):[7]

Mandarin Chinese (Sino-Tibetan) (Huang & Ahrens 2003: 361)

(5) a. *yi* *duo* *hua*
one CLF flower(s)/plant(s)
'one flower'

b. *yi* *zhu* *hua*
one CLF flower(s)/plant(s)
'one plant'

C17P27 Seen from a European perspective, one could say that sort nouns such as Mandarin *hua*$_{N}$ 'flower(s)/plant(s)' are semantically vague or underspecified and have a wider meaning than a singular object noun like English *plant*$_{N}$, which can be said to have a narrow meaning in comparison. Nevertheless, it is important to emphasize that singular object nouns and sort nouns are equally good exemplars of first order nouns and the same holds true for members of other nominal subcategories that are used to talk about spatial objects in the real world (Rijkhoff 2008: 738), such as set nouns and general nouns, which are discussed below.

[7] Sometimes we also find variable classification in languages with a genuine gender or noun class system, in which nouns normally always belong to the same gender or noun class (Dixon 1986). For example, the Danish noun *studie*$_{N}$ occurs with common and neuter gender, but there is a difference in meaning: *en studie*, with common gender, means 'a branch of learning' (e.g. architecture or mathematics), whereas *et studie*, with neuter gender, means 'a thesis'. The Danish noun *hamster*$_{N}$ 'hamster' can also be used with common or neuter gender (*en hamster* or *et hamster*), but in this case there is no semantic difference. Here gender shift is due to the speaker's individual preference or incomplete language change (Trap-Jensen 2007; Skafte-Jensen 2011; see Visser 2011 on historical gender change in West Frisian). Since the lexicalized meaning is not vague or ambiguous in these Danish examples, they are not cases of semantic under-specification. On the special case of polarity, the reversal of gender between the singular and the plural of nouns, see, e.g. Reh (1985) & Nilsson (2016); also Rijkhoff 2004: 85–86).

C17P28 Sort nouns like Thai *mánaaw*$_N$ 'lemon(s)', which combine with a sortal classifier, should not be confused with mass nouns like *water*$_N$ or *oil*$_N$ (section 17.3.3.3), which occur in combination with a measure word or a mensural classifier as in *a litre/bucket/bottle/cup* (*of water*); on grammatical and other differences between sortal and mensural classifiers, see e.g. Seifart (2010: 722).

C17P29 In sum, it appears that a positive, negative, or neutral value for the semantic features Shape and Homogeneity can be used to characterize first order nouns in languages across the globe (Rijkhoff 2004: 28–59; Rijkhoff 2008; the reader is referred to Part I of this volume for overviews of various theoretical approaches to word classes). A classification of nouns based on these features is shown in Figure 17.1, in which each nominal subtype is lexically specified for a different *Seinsart* (literally 'mode of being', the nominal counterpart of the established German term *Aktionsart* 'mode of action' in verb semantics; see Rijkhoff 2004: ch. 2, for details).

	–HOMOGENEITY	+HOMOGENEITY
+SHAPE	*Singular Object Noun*	*Collective Noun*
	Set Noun	
–SHAPE	*Sort Noun*	*Mass Noun*
	General Noun	

C17F1 **FIGURE 17.1** Nominal subcategories ('*Seinsarten*'; based on Rijkhoff 2004: 54)

C17P30 Each of the nominal subtypes in Figure 17.1 is discussed below, including mass nouns, even though they are not normally used to talk about concrete objects in the physical world (but see sections 17.4.2–17.4.3). Section 17.3.2 is concerned with nouns that have a positive value for the feature Shape as part of their lexicalized meaning, which makes it possible for these nouns to be counted directly (hence: count nouns): singular object nouns, collective nouns, and set nouns. Section 17.3.3 discusses nouns that have a negative value for the feature Shape, which is the reason they cannot be counted directly: sort nouns, general nouns, and mass nouns. Notice that several nominal subtypes in Figure 17.1 may co-occur in the same language. For example, in languages that have singular object nouns (section 17.3.2.1), they may co-occur with collective nouns (section 17.3.2.2), set nouns (section 17.3.2.3), or mass nouns (section 17.3.3.3).

C17P31 If we temporarily ignore collective nouns and mass nouns, i.e. nouns with a positive value for the feature Homogeneity (+Homogeneity), it appears that across languages four kinds of nouns are used to talk about the same single object in the real world (like a banana or an umbrella): singular object nouns, set nouns, sort nouns and general nouns.

C17P32 It is important to underscore, once again (see the quotation above from Levin & Rappaport Hovav 2011), that Shape and Homogeneity in Figure 17.1 do not refer to properties of real-world objects or substances, but rather specify coded values for semantic features in the lexicalized meaning of a first order noun. It should also be stressed that the nominal subcategories in Figure 17.1 are idealizations, as boundaries between linguistic categories are not always clear cut (see also section 17.4).[8] One of the reasons for this is that categorization is sometimes a matter of (inadequate) perception (Russell 1923). Thus 'hair' is a mass noun in 'long hair' and a singular object noun in 'a long hair'. Language change is another reason that

[8] On gradience and continuity in grammar, see, e.g. Bolinger 1961; Sasse 1993; Aarts 2007; Aarts et al. 2004; Sorace & Keller 2005; Langacker 2006: 116.

nouns may not always fit neatly into one of the categories of Figure 17.1. Linguistic categories (just like languages themselves) are dynamic entities and since processes of language change do not happen overnight, this means we may also encounter nouns whose membership of a specific subcategory is (currently) not quite clear (sections 17.3.3.1 and 17.4.1).

C17P33 It is not obvious why the feature Shape (relevant for all six nominal subcategories in Figure 17.1) would be more basic or significant than the feature Homogeneity in noun semantics or in grammar generally. Still, the classification in Figure 17.1 confirms Friedrich's (1970: 380) observation that 'the category of shape appears to be a typological universal in grammar […] and of not inconsiderable significance for a theory of semantics in grammar' (see Wierzbicka 2006, Rijkhoff 2008; see also section 17.3.4). Friedrich (1979: 256) further demonstrated the widespread influence of Shape in different areas of grammar in languages across the globe, suggesting that 'shape ranks with time, aspect and the like as one of the fundamental ideas in grammar and, by implication, in the mind of the speakers of natural language' (see also e.g. Mithun 1999: 104–117).

C17P34 One reason why the feature Shape takes precedence over the feature Homogeneity may be due to the idea that spatial orientation is primary in human cognition, which links up well with observations about the fundamental role of spatial notions in various grammatical domains, including lexical semantics (see e.g. Gentner 2001; Goddard 2001; Casasanto & Boroditsky 2008; Majid et al. 2013), sometimes discussed under the heading 'localism' (Lyons 1977: 718–724; Fortis 2018). The role of shape in grammar and cognition has often been investigated by comparing languages using different types of nouns for the same object in the real world; more on this topic in section 17.3.4.

C17P35 Notice, finally, that since the classification in Figure 17.1 concerns nouns in their basic, unmarked form, plural nouns do not constitute a nominal category by themselves in the current approach, as opposed to certain logical treatments of nouns, where plural nouns are often discussed together with collective or mass nouns (Lasersohn 2011).

C17S5 17.3.2 Nouns that can be in a direct construction with a cardinal numeral (count nouns)

C17P36 Figure 17.1 shows that there are three types of noun whose meaning includes a positive value for the feature Shape, which means that they can be counted directly. The three + Shape nominal subcategories can be characterized as follows:

C17P37 • SINGULAR OBJECT NOUNS (+Shape, −Homogeneity) like English *chair* or *cat* designate a non-agglomerative, spatially bounded property of a single object;

C17P38 • COLLECTIVE NOUNS (+Shape, +Homogeneity) like English *crew* or *swarm* designate an agglomerative, spatially bounded property that applies to all the members of what is regarded as a group;

C17P39 • SET NOUNS (+Shape) like Lango *gúlú* 'pot(s)' or Turkish *çocuk* 'child(ren)' designate a spatially bounded property of one or more objects.

C17P40 The fact that nouns with a positive value for the feature Shape are counted directly (hence the label 'count nouns') does not necessarily mean that they can be marked for plural number (Allan 1980; Acquaviva 2008). This is only true in the case of singular object

nouns (cat_N—$cat\text{-}s_N$) and collective nouns ($swarm_N$—$swarm\text{-}s_N$), both with and without a modifying numeral (on the various kinds of plurals, see Corbett 2000).[9] This is shown in examples (6) and (7) from Dutch (under certain circumstances a plural marked form can have a number-neutral interpretation; Sauerland et al. 2005). An example of a set noun, which is transnumeral, is given in (8); set nouns are discussed in more detail in section 17.3.2.3. In each case (singular object noun, collective noun, set noun), the noun can be directly modified by a cardinal numeral; hence they are assumed to be lexically coded for the feature + Shape. Notice that Lango nouns (or rather: noun phrases) are usually not marked as definite or indefinite (Noonan 1992: 161).

Dutch: SINGULAR OBJECT NOUN

(6) a. *een* *boek*
a book
'a book'

b. *boek-en*
book-PL
'books'

c. *twee* *boek-en*
two book-PL
'two books'

Dutch: COLLECTIVE NOUN

(7) a. *een* *kudde*
a herd
'a herd'

b. *kudde-s*
herd-PL
'herds'

c. *twee* *kudde-s*
two herd-PL
'two herds'

Lango (Nilotic) (Noonan 1992: 167): SET NOUN (a set noun is transnumeral)

(8) a. *gúlú*
pot(s)
'a/the pot(s)'

b. *gúlú* *àryô*
pot(s) two
'(the) two pots'

C17P41 In some languages, nouns are always overtly marked for number, possibly also for e.g. gender or case (see Armoskaite 2019 on Lithuanian, where no bare roots are allowed). Below are some examples from Sesotho, a representative of the large Bantu family. In Bantu languages, the singular–plural opposition is expressed through the noun class system by (presumably) derivational affixes (on the derivational nature of noun class markers, see e.g. Schadeberg 2001 and references in Cobbinah & Lüpke 2014: 206). Generally speaking, countable nouns are members of two noun classes, one singular and one plural (the number before the Sesotho noun indicates noun class, e.g. *thọ* 'person(s)' belongs to Noun Class 1 in the singular and Noun Class 2 in the plural).

Sesotho (Bantu): singular–plural opposition in some noun class pairings (Demuth et al. 1986: 455)

(9)

CLASS	SINGULAR		CLASS	PLURAL	
1	*mọ-thọ*	'person'	2	*ba-thọ*	'persons'
3	*mọ-sẹ́*	'dress'	4	*mẹ -sẹ́*	'dresses
5	*lẹ-tsatsí*	'day'	6	*ma-tsatsí*	'days'
7	*sẹ-liba*	'spring/well'	8	*li-liba*	'springs/wells'
9	*Ø-ntjá*	'dog'	10	*li-ntjá*	'dogs'

[9] Recall that the classification in Figure 17.1 concerns nouns in their basic, unmarked form, which means that plural nouns do not constitute a nominal category in their own right here (section 17.3.1).

C17P42 Apparently, some languages from the Atlantic branch of Niger–Congo, such as Banyun or Baïnounk, distinguish between two kinds of plural: the 'limited/count plural' and the 'unlimited/collective plural' (see Greenberg 1972: 20, fn.23, quoting Sauvageot 1967). Further details on the so-called unlimited plural in three Baïnounk languages can be found in Cobbinah & Lüpke (2014) and in Corbett (2000: 30–35, 238), who discusses it under the heading 'greater plural' or 'plural of abundance'.

C17P43 In Zuni (a language isolate spoken in New Mexico) nouns are also overtly marked as singular or plural, according to Newman (1965: 56; as quoted in Mithun 1999: 79).

Zuni (Language Isolate): inflectional number on Zuni nouns

(10) a. *ča-ʔleʔ* 'child'
b. *ča-weʔ* 'children'

C17P44 Presumably languages like Sesotho & Zuni have transnumeral nominal roots, but since nouns do not appear in their unmarked form, it is difficult to determine what kind of nominal subcategory they belong to. Therefore, such 'bound nouns' are ignored in the current chapter (see also Mithun 1999: 57 on the relation between bound morphology and lexical categories). In languages with a case system, the nominative singular is typically used as the reference form of the noun, as this is usually regarded as the unmarked case. A related problem determining nominal subcategories concerns languages in which number marking is expressed outside the noun phrase proper, commonly as an optional phrase-final clitic which is clearly related to the third person pronoun or as an optional appositional element that is mutually exclusive with a numeral. This is the case in, for example, Alamblak, Bambara, and Nama Hottentot (Rijkhoff 2004: 31–32). The optional character of these markers suggests that the nouns are transnumeral set nouns (section ~~17.3.3~~), but since sufficient data are not available, such cases will also remain undiscussed here. Below is an example from Guaraní, where the (so-called) plural marker appears after all postnominal modifiers and is mutually exclusive with numerals (Gregores & Suárez 1967: 144, 150, 155).

Guaraní (Gregores & Suárez 1976: 150)

(11) *upé la ʔógoità nte gwigwà kwéra*
that the house(s) just stone of PL
'just those houses of stone'

C17S6 ### *17.3.2.1 Singular object nouns*

C17P45 As shown in (6), a singular object noun denotes a property of single object (person, animal, plant, thing) and is marked for plural number if reference is made to more than one object, both with and without a modifying cardinal numeral higher than 'one':

Dutch: SINGULAR OBJECT NOUN

(12) a. (*een*) *tafel*
(a) table
'(a single) table'
b. *tafel-s*
table-PL
'(multiple) tables'
c. *twee tafel-s*
two table-PL
'two tables'

C17P46 In Dutch, the plural suffix is normally realized as *-en* or *-s*, but some nouns have an irregular plural. For example, the plural form of *kind* 'child' is *kinderen* 'children', which is an instance

of metanalysis (or 'rebracketing'): when at some point in the history of Dutch the original plural suffix *-er* on *kind* 'child' was no longer recognized as a pluralizing morpheme, the plural suffix *-en* was added to the erstwhile plural form *kinder*, yielding *kinderen* 'children' (see Harris 2017 for a comprehensive treatment of MULTIPLE EXPONENCE, the occurrence of multiple realizations of a single morphosemantic feature). Diachronic developments also explain why English has so-called invariant plurals, which seem mostly restricted to animal-denoting nouns such as (*one*/*two*/*many*) *deer*, *fish*, or *sheep*. Irregular plural forms of English nouns can also be the result of a vowel shift (*Umlaut*), as in singular *foot* vs plural *feet*.

C17P47 When the plural form involves a different root or stem, we speak of total suppletion. In Lango, for example, only some nouns have a plural form (mainly those denoting humans and animals), but since there is no regular plural formation 'each singular/plural pair must be individually lexicalised' (Noonan 1992: 167), as in *dákô* 'woman' vs *món* 'women'. Most Lango nouns, however, are transnumeral set nouns (section 17.3.2.3). Apparently lexical plurals are typically attested in languages with 'limited plural marking' (Mithun 1999: 83). In addition to suppletion, Ngiti (a Nilo-Saharan language) has two other strategies of plural marking: the use of Bantu-like prefixes *mU-* and *pba-*, and tone (Kutsch Lojenga 1994: 133f.).

C17P48 Plural number (if it is that) can also be expressed through partial or full reduplication, a morphological process that has many functions, including marking distributivity or collectivity (Moravcsik 1978a; Rubino 2005). In addition to the plural, languages may have what has been referred to as 'minor number categories', such as the dual, the trial or the paucal (see Universal 34 in Greenberg 1966: 94; Corbett 2000; Harbour 2020).

C17P49 Notice finally that in a few languages the same so-called number marker can have different values (see Tiersma 1982 on local markedness). This is attested in languages of the North American Kiowa–Tanoan family, which have only the suffix *-sh* to mark the unexpected or inverse number (Mithun 1999: 81; see also Grimm (2012) on Dagaare, a Gur language of the Niger–Congo family). For example, the unmarked number of animate nouns is singular, so on these nouns *-sh* marks dual or plural number, but in the case of inanimate nouns, which tend to be used in connection with less individuated or multiple entities, plural is the expected number, so here *-sh* marks singular or dual number.

C17S7 *17.3.2.2 Collective nouns*

C17P50 A collective noun denotes a property of objects (persons, animals, plants, things) which are conceived of as a unit and which is marked for plural number if reference is made to more than one collective entity, both with and without a modifying cardinal numeral higher than *one*. The fact that these nouns can be counted directly suggests they have a positive value for the feature Shape, just like singular object nouns (section 17.3.2.1) and set nouns (section 17.3.2.3).

English (Germanic): COLLECTIVE NOUN

(13)	a.	(*a*)	*crew*	b.	*crew-s*	c.	two	*crew-s*
		(a)	crew		crew-PL		two	crew-PL
		'(a single) crew'			'(multiple) crews'		'two crews'	

C17P51 Greenberg (1972: 25) already observed that the term COLLECTIVE NOUN has been given various interpretations, which has led to confusing analyses (see also Gil 1996; Nurmio

2017: 66). For example, Wiese (2012: 55) does not regard group nouns like *family* or indeed *group* as collective nouns. Here we will confine ourselves to collective nouns in the strict sense of the word: nouns which can be counted directly, which can be pluralized and which in their basic, unmarked form are used to refer to a group of people, animals, or objects like English *herd*, *team* or *family* (see Kuhn 1982; de Vries 2021).

C17P52 Collective nouns (like mass nouns) have a positive value for the feature Homogeneity, because they are cumulative (or agglomerative) and dissective (or divisive) entities (Goodman 1966; Rijkhoff 2008: 734–735; Doetjes 2012: 2561). When someone in a family dies, the others are still family, and when we take one flower out of a bunch, the remaining flowers are still part of a bunch—up to a point (Quine 1960: 91). Since the property of dissectiveness or divisibility has a lower limit (do two flowers together constitute a bunch?), it is more attractive to say that homogeneous entities are agglomerative or cumulative in that they can be expanded with more members or portions without affecting the grammatical number. By adding members to a collective we do not get more collectives, as it will only increase the size of a collective, and by adding portions to a mass, we only increase the size, weight, or volume of that mass entity. In other words, collective nouns and mass nouns can be said to have 'cumulative reference' (but strictly speaking it is noun phrases rather than nouns that have referential potential). For instance, when a child is born into the family, this does not result in more families; it only affects the number of individuals that make up that family. Notice that the number of individuals in certain collectives, such as soccer teams, is fixed by the rules of the game.

C17P53 There is considerable variability within and across varieties of English regarding agreement patterns involving subject noun phrases headed by a collective noun (Depraetere 2003; Hundt 2009). In British English, the focus can be on the group or on the members in that group, hence we find singular as well as plural verb agreement, whereas in American English collective nouns usually trigger singular verb agreement (Corbett 2006: 206–213). Collective nouns often denote groups of animals (e.g. *clowder*$_{N}$ *of cats*), and the same is true for (British) English, where many such nouns are used in connection with hunting events (for an early list, see Berners [1486] 1881; cf. Biber et al. (1999) on 'quantifying collectives'), like *pride*$_{N}$ (*of lions*), *gaggle*$_{N}$ (*of geese*), or *parliament*$_{N}$ (*of owls*).

C17P54 In addition to these lexical collectives, some languages have what are sometimes called morphological collectives. Like lexical collectives, they are used to refer to units containing multiple members, but they can be affixed with a singulative marker to denote a single object. Compare, for example, Breton *gwez* 'trees, forest' (collective) and *gwez-enn* 'a (single) tree'. In its turn, *gwez-enn* can be inflected with a regular plural suffix: *gwezenn-où* '(multiple individual) trees'. Whereas the so-called lexical collectives in Welsh, another Brittonic language, often denote human groups, members of the relatively small group of morphological collectives are typically used to talk about non-human collectives, like groups of animals, insects, fruits, or vegetables; compare the Welsh collective *moch* 'pigs' and *moch-yn* 'a (single) pig' (Stolz 2001; Nurmio 2015 and 2017).[10] There are, however, good reasons to assume that the singulative in, for example, Breton is a derivational rather than an inflectional suffix,

[10] The discussion of lexical and morphological collectives in Brittonic languages has a complicated history, with various conflicting analyses. It should also be added here that in Welsh, for example, the so-called lexical collective or corporate noun (Corbett 2000: 188) does not appear to have the same properties as English collective nouns like *swarm*$_{N}$ or *crew*$_{N}$. For a detailed discussion of the two kinds of 'collectives' in Welsh, see Nurmio (2017).

since it also serves as a nominalizing morpheme on verbal and adjectival bases (Greenberg 1972: 20; Acquaviva 2008: 246). Derivation is also used to turn mass nouns such as Dutch *bier* 'beer' into words denoting a single countable unit, as in *een bier-tje* (a beer-DIM) 'a glass of beer' (Jurafsky 1996).

C17P55 As shown in section 17.3.2.3, set nouns can also appear with a singulative marker (which is presumably inflectional rather than derivational), but there is an important difference: whereas an unmarked morphological collective noun always denotes a group of multiple objects (like Breton *gwez* 'trees, forest'), an unmarked set noun is transnumeral, which means it can also be used to talk about a single object.

C17P56 In numeral classifier languages like Thai (examples (3) and (4); see also section 17.3.3.1), expressions that are used to talk about collective entities contain what is sometimes called a (quantitative) collective classifier, as in this example from Burmese (Okell 1969: 211): *pyà hnă ouñ* (bee(s) two swarm$_{CLF}$) 'two swarms of bees'. Collective classifiers (as well as collective aspect markers, section 17.3.2.3) constitute one of the historical sources from which plural markers may develop (Rijkhoff 2004: 116–117). Collective nouns are assigned their own noun class (often noun class 14) in Bantu languages (section 17.3.2). For an account of the way speakers talk about collective entities like a choir or a boat crew in Samoan, a language that is claimed not to distinguish syntactically between major lexical word classes, see Mosel & Hovdhaugen (1992: 89–93).

C17S8 ### *17.3.2.3 Set nouns*

C17P57 The third nominal subcategory whose members can be directly modified by a cardinal numeral concerns the number-neutral or transnumeral SET NOUNS, so-called because they designate a set entity, rather than an individual object or a collective. A set may contain any number of individuals, including 'one' (in which case we talk about a singleton set). Set nouns are attested in many languages, where they often occur alongside members of another nominal subcategory, such as singular object nouns (section ~~17.3.1~~; see also sections ~~17.3.3.1~~ and 17.4.1 on set nouns co-occurring with sort nouns). In such cases, the singular object nouns tend to be used to talk about humans or higher animals, whereas the set nouns usually denote lower animals or inanimate entities (Rijkhoff 2004: 30).

C17P58 Since set nouns have only recently been recognized as a distinct nominal subcategory, they will receive some extra attention here. Set nouns are sometimes called general number nouns (Corbett 2000: 9–18, 73), together with sort nouns (section 17.3.3.1), but there are important differences (n. 6); for example, set nouns—unlike sort nouns—do not require a numeral classifier when modified by a cardinal numeral. Confusingly, set nouns are sometimes regarded as collective nouns or mass nouns (e.g. Ojeda 1998: 249; Corbett 2000: 5, 13; Wiese 2012).

C17S9 #### *17.3.2.3.1 Set nouns and numerals*

C17P59 When a set noun is modified by a numeral, it remains in its unmarked, transnumeral form, as shown in these examples from Oromo and Turkish (see Lewis 1967: 25–26; see Schroeder 1999 for a detailed study of Turkish nouns and noun phrases):

Oromo (Cushitic) (Stroomer 1987: 107)

(14)	a.	*gaala*	b.	*~~gaal~~*	*lamaani*
		camel(s)		camel(s)	two
		'camel' or 'camels'		'two camels'	

Turkish (Altaic) (Görgülü 2012: 69; Bamyacı et al. 2014: ~~25~~)

(15) a. *çocuk*
child(ren)
'child' or 'children'

b. *beş* *çocuk*
five child(ren)
'five children'

AQ: Please provide reference details for Görgülü 2012.
DONE

C17P60 This suggests that a cardinal numeral which modifies a set noun serves a different function than a numeral that modifies a member of the other two + Shape noun types. Whereas in the case of a singular object noun or a collective noun the modifying numeral seems to multiply the number of singular objects or collectives (hence the plural *-s* suffix in *two umbrella-s* and *two crew-s*), in the case of a set noun the numeral seems to specify the number of members that make up the set. In other words, when a cardinal numeral modifies a set noun, we do not get multiple sets; instead the numeral seems to indicate the size of the (single) set.

C17P61 In some languages, a noun phrase headed by a numerated set noun triggers singular number agreement on the verb. This has been labelled 'number discord', which appears to be a misnomer because number agreement is with the set, which is always a singular entity, whether it contains one or more individuals. Below is an example from Oromo, but the same phenomenon is attested in other languages with subject–verb agreement in which the subject noun phrase is headed by a set noun, such as the Caucasian (Kartvelian) language Georgian (Harris 1981: 22) and Lango (Noonan 1992: 168).

Oromo (Stroomer 1987: 59, 107)

(16) *gaala* *lamaani* *sookoo* *d'ak'-e*
camel(s) two market go-3SG.M.PAST
'Two camels went to the market.'

C17S10 *17.3.2.3.2 Set nouns and nominal aspect marking*

C17P62 In some languages, set nouns can optionally appear with what superficially looks like a plural number marker (although this element is normally absent when the set noun is modified by a cardinal numeral) or even a so-called singulative number marker (see e.g. Stroomer 1987: 74–88). However, due to the distinctive semantic and formal properties of these so-called number markers, it has been argued that they are better analysed as members of another grammatical category: NOMINAL ASPECT MARKERS (the nominal counterparts of verbal aspect markers; Rijkhoff 2004: 101–121). Whereas number markers are concerned with quantitative properties of the referent, nominal aspect markers specify a QUALITATIVE property of the referent, viz. the KIND of set the speaker is referring to: a set with just one member (singleton set) or a set that has multiple members (collective or distributive set). Confusingly, the term 'singulative marker' is also used in connection with morphological collectives (section 17.3.2.2), but whereas this element seems to derive a singular object noun from a morphological collective noun (and collectives always involve multiple individuals), a singulative marker on a transnumeral set noun has a specifying (disambiguating) function.

C17P63 Notice that genuine number marking would be logically or semantically incompatible with the transnumeral character of set nouns: the 'singular' or 'plural' of a number-neutral noun leads to nonsensical analyses (see also section 17.3.3.1 on the absence of number marking on sort nouns, another transnumeral subtype). Rather than indicating the number of individual objects or collectives (sections 17.3.2.1 and 17.3.2.2), nominal aspect markers

specify whether the speaker refers to a singleton set (with just one member) or to a collective or distributive set (with multiple members). For example, in Oromo the collective marker is used on set nouns to refer to a group of individuals or objects (Stroomer 1987: 76). Nouns in several North American indigenous languages may occur with a distributive marker, which indicates the spread of individuals or objects of a group over various locations ('here and there') or over various sorts (Boas 1911a: 37–38; Mathiot 1967; Ojeda 1998: 248f.; Mithun 1999: 88–93; Corbett 2000: 111–120). For example, Mohawk *onén:ia'* means 'rock(s)' and *onenia'-shòn:'a* 'various rocks' (of assorted types, usually consisting of different shapes, sizes, and colours). In many languages, there is a historical connection between collective or distributive markers and plural markers. For example, a diachronic development from collective or distributive to plural has been observed in languages of Native North America (Boas 1911b: 444; Mithun 1999: 91) as well as languages from the Kartvelian, Mesoamerican, and Semitic families (Rijkhoff 2004: 116–117, fn. 22). Such a diachronic change also implies a shift from set noun to singular object noun (section 17.4.1).

C17P64 It should be noted, finally, that our understanding of expressions of nominal collectivity and distributivity in languages across the globe is still rather inadequate, which is at least partly due the fact that the notions collectivity and distributivity, both of which presuppose multiplicity, have been defined and applied in rather different ways (Gill 1996; Corbett 2000: 111, 117). Notice also that in some languages a collective or distributive marker can (optionally) co-occur with a (true or apparent) plural marker, which is why, for example, Corbett (2000: 120) prefers to regard these markers as members of categories that are 'distinct from but related to number'.

C17S11

17.3.3 Nouns that cannot be in a direct construction with a cardinal numeral

C17P65 Three nominal subtypes have a negative value for the feature Shape (−Shape):

C17P66
- SORT NOUNS (−Shape, −Homogeneity) like Thai *rôm*$_{N}$ 'umbrella(s) or *mánaaw*$_{N}$ 'lemon(s)' designate a non-agglomerative, spatially unbounded property of one or more objects;

C17P67
- GENERAL NOUNS (−Shape) like Yucatec Mayan *háʼas* (see (19) for examples) designate a property that is only characterized as not having a definite spatial outline;

C17P68
- MASS NOUNS (−Shape, +Homogeneity) like English *water* or *butter* designate an agglomerative property that is characterized as not having a definite spatial outline.

C17P69 Members of these nominal subtypes cannot be in a direct construction with a cardinal numeral; modification by a numeral typically requires the employment of an individualizing element such as a sortal or mensural classifier (languages in which nouns that are used to talk about masses can combine directly with a numeral are discussed in section 17.4.3).

C17P70 The role of shape in grammar and cognition has often been investigated by comparing languages using –Shape nouns (notably 'classifier languages') with languages using + Shape nouns (Foley 1997: 239–245; also e.g. Lucy 1992 and Evans 2010). For example, Perniss et al. (2012) is concerned with the role of shape in Bora (a classifier language spoken in the

Amazonian regions of Columbia and Peru) and they found that its speakers systematically encode shape, more so than speakers of English and Spanish. Lucy (1992) is a comparative study of relations between language and cognition in English and Yucatec Maya, another classifier language (see section 17.3.3.2). Native speakers of both languages were first shown a single object and then two different objects. Subsequently they had to say which of the two different objects they believed resembled the first object most. It turned out that native speakers of English usually choose the object on the basis of its shape (e.g. box-shaped). Yucatec speakers, on the other hand, commonly focused on the kind of material the object was made of (e.g. cardboard), suggesting that the shape of the object played a far less important role (see Bowerman & Choi 2003 and Bowerman (2018: 134) on the role of first language acquisition in spatial categorization). Imai & Gentner (1993; 1997) reported similar results when they compared speakers of English and Japanese, another language that is deemed to use –Shape nouns for concrete objects (section 17.3.3.1). Li et al. (2009) is an attempt to account for these facts (which seem to confirm the idea of linguistic relativity, also known as the 'weak Whorfian hypothesis') in non-relativistic terms.[11]

C17P71 The three –Shape subtypes (sort nouns, general nouns, mass nouns) are discussed in turn below.

C17S12 ### *17.3.3.1 Sort nouns*

C17P72 Members of the nominal subcategory SORT NOUN are used to refer to one or more concrete objects in the real world, even though these nouns do not seem to include the notion of spatial boundedness or discreteness as part of their lexical meaning (Lyons 1977: 460–466; Hundius & Kölver 1983; Bisang 1999).[12] As mentioned earlier, sort nouns as attested in, for example, languages spoken in East and mainland South East Asia are not marked for number and generally need individualizing elements, called sortal classifiers (a.k.a. numeral or count classifiers), when they are modified by a cardinal numeral.[13]

AQ: Please clarify whether Craig (1996) should be 1986 in Note 13.
YES: 1986

[11] On linguistic relativity, see, e.g. Grace 1987; Gumperz & Levinson 1996; Boroditsky 2003; Levinson 2003b; LaPolla 2020.

[12] In logical approaches to noun semantics, sort nouns have been analysed as members of various traditional, European nominal subcategories, such as collective nouns (Wiese 2012: 55) or mass nouns (Chierchia 1998b).

[13] Confusingly, the term 'classifier' has been used for a variety of elements; see, for example, Schroeder's (1999: ch. 4) discussion of 'the "so-called classifier" *tane*' in Turkish. The sortal (numeral, count) classifiers discussed in this section are typically attested in languages of East and mainland South-East Asia and must appear when sort nouns are counted. Elsewhere, in particular in e.g. Mesoamerica (e.g. Jacaltec) and in the Amazon region (e.g. Tariana or Yagua), elements that have also been called numeral classifiers are typically morphologically bound forms which also have inflectional or derivational properties (Payne 1987; Derbyshire & Payne 1990). Furthermore, these forms are often all part of a larger system of classificatory elements in the noun phrase, in which a single, complex system of 'multiple classifiers' seems to be composed of various subtypes (Craig ~~1996~~; Aikhenvald 2000; Seifart 2010: 722, 724). Systems with multiple or mixed classifiers are ignored here, because they tend to become 'increasingly heterogeneous and typological predictions associated with them (e.g. type-specific semantics or grammaticalization patterns) are weakened' (Seifart 2010: 724). As noted below, in some languages (erstwhile) numeral classifiers have also assumed non-individualizing functions due to grammaticalization (Bisang 2008c).

Thai (Hundius & Kölver 1983: 167)

(17) *lûuk* *săam* *khon*
child(ren) three CLF:PERSON
'three children'

C17P73 The wide meaning of sort nouns is often counterbalanced through coercion (Huang & Ahrens 2003) or by pragmatic factors (see Bisang 2009 and 2014 on hidden complexity). It is perhaps good to recall (section 17.3.1) that sort nouns as well as members of the other nominal subcategories discussed here are equally good exemplars of the lexical category Noun from a cross-linguistic perspective (Rijkhoff 2008: 738–739). The ambiguous or vague semantics of nouns with a 'wide meaning' (set nouns, sort nouns and general nouns) may be difficult to handle for grammatical theories that do not take contextual factors into account, but it makes them rather versatile ('flexible') for referential purposes (see also section 17.4.3 on the role of context).

C17P74 As an illustration of the communicative or pragmatic versatility of sort nouns, consider these well-known examples from Burmese (Sino-Tibetan), which has various classifiers for *myiʔ* 'river(s)'. According to Becker (1975: 113; see also Hle Pe 1965; Vittrant 2005), it is possible to refer to a river in at least eight contexts:

Burmese (Sino-Tibetan) (Becker 1975: 113)

(18) *myiʔ tə yaʔ* 'river one place' (e.g. destination for a picnic)
myiʔ tə tan 'river one line' (e.g. on a map)
myiʔ tə hmwa 'river one section' (e.g. a fishing area)
myiʔ tə ʻsin 'river one distant arc' (e.g. a path to the sea)
myiʔ tə θwɛ 'river one connection' (e.g. tying two villages)
myiʔ tə ʻpa 'river one sacred object' (e.g., in mythology)
myiʔ tə khuʼ 'river one conceptual unit' (e.g., in a discussion of rivers in general)
myiʔ tə myiʔ 'river one river' (the unmarked case)

C17P75 The number of sortal classifiers differs from language to language. If a language employs a large set of numeral classifiers, the way they are used depends on the speaker's social status and competence (Adams 1989). For a more detailed discussion of numeral classifiers, sometimes in the larger context of 'nominal classification' or 'noun categorization devices', see e.g. Aikhenvald 2000, Grinevald 2000 and 2007; Rijkhoff (2004: 115–117, 162–166); Seifart 2010; Bisang & Wu 2017.

C17P76 It has been known for some time that sortal classifiers and plural markers appear to be mutually exclusive (Greenberg 1972; Sanches & Slobin 1973; it was stated in section 17.3.2.3 that genuine number markers are semantically incompatible with transnumeral nouns). Recently it has been suggested that there may be exceptions to this claim and that the so-called Greenberg–Sanches–Slobin generalization is a statistically highly significant correlation rather than an absolute universal (Doetjes 2012; Tang & Her 2019; see Bisang 2012 and other chapters in Xu 2012 for critical discussion). It should be noted, however, that at least some of the (apparent) counterexamples involve an optional collective aspect marker on a set noun (section 17.3.2.3) rather than a true, compulsory plural marker on a singular object noun. For example, the so-called plural suffix *-men* in Mandarin Chinese is optional,

incompatible with counting and restricted to human nouns, as in *háizi-men* '(a certain group of) children' (Norman 1988: 159). Iljic (2001: 95; see also Iljic 2005) calls *-men* a 'personal collective' which 'differs radically from the Indo-European plural endings' (~~ibid. 74~~) and argues that *-men* is probably the result of the grammaticalization of the Mandarin translational equivalent of 'clan'. This could indicate that certain nouns in Mandarin Chinese are in the process of changing (or already have changed) from sort noun to set noun.

AQ: Please replace ibid. with relevant ref.

C17P77 It is important to note that due to language change, (erstwhile) numeral classifiers can assume non-individualizing functions in combination with e.g. demonstratives or adjectives (or rather stative verbs, which are typically used to express adjectival notions in South East Asian languages), where their appearance is often optional rather than compulsory (McGregor & Wichmann 2018). For example, Bisang (2008c) shows that Thai and other languages of East and mainland South East Asia also have highly grammaticalized classifiers, which no longer serve as individualizers of sort nouns, but have preserved their original phonological shape (i.e. grammaticalization without coevolution of form and meaning).

C17S13 ### *17.3.3.2 General nouns*

C17P78 There are also languages with transnumeral nouns that are used to talk about both objects and masses, called 'general nouns' (Rijkhoff 2004: 49f.). There do not seem to be many languages employing such general nouns, but Yucatec Maya appears to be one of them. According to Lucy (1992: 76, 83), this language does not make a fundamental distinction between sortal classifiers (section ~~17.3.1~~) and mensural classifiers (section ~~17.3.3~~) in the case of certain nouns (which may be few in number): hence we have labelled these elements 'general classifiers'. Here are some examples of general classifiers which all involve the general noun *há'as* and which, as shown below, are rather difficult to translate into English.

Yucatec Maya (Mayan) (Lucy 1992: 74)

(19)	a.	*'un-ɛ'ɨit*	*há'as*	'one/a 1-dimensional banana (i.e. the fruit)'
	b.	*'un-wáal*	*há'as*	'one/a 2-dimensional banana (i.e. the leaf)'
	c.	*'un-p'ɛ̈el*	*há'as*	'one/a 3-dimensional banana (i.e. the fruit)'
	d.	*'un-kúul*	*há'as*	'one/a planted banana (i.e. the plant/tree)'
	e.	*'un-kúuch*	*há'as*	'one/a load banana (i.e. the bunch)'
	f.	*'um-p'ɨit*	*há'as*	'a_little_bit_of/some banana'
		a/one-CLF	banana(s)	

C17S14 ### *17.3.3.3 Mass nouns*

C17P79 Just like 'collective noun', the label 'mass noun' is sometimes applied in rather unusual ways. For example, in certain logical approaches to word classes (e.g. Borer 2005a: 108), it has been suggested that at some abstract, underlying level of representation all nouns start out as mass nouns. By contrast, this section is concerned with mass nouns as an overt nominal subcategory, i.e. nouns whose lexicalized meanings have a negative value for the feature Shape (−Shape) and a positive value for the feature Homogeneity (+Homogeneity). As indicated in section 17.3.1, this does not necessarily mean that there is a direct relationship between the denotation of a mass noun and what counts as a mass from a physical or ontological standpoint. This is shown by the fact that the same physical entity can be referred to

by using a count noun in one language and by a mass noun in another language, as in the case of English *onion*$_{\text{CountN}}$ and its Russian translational counterpart *luk*$_{\text{MassN}}$ (Wierzbicka 1985: 314): the difference between mass nouns and count nouns rests in the coded features of the nouns rather than properties of physical entities in the external world.

C17P80 The current chapter does not explore possible cognitive or cultural reasons why, within or across languages, speakers are not consistent in discriminating between objects and masses (Koptjevskaja-Tamm 2006). Thus, any motivations behind the choice between mass noun vs count noun in cases like 'foliage'$_{\text{Mass}}$ vs 'leave(s)'$_{\text{Count}}$, 'gravel'$_{\text{Mass}}$ vs 'pebble(s)'$_{\text{Count}}$ or 'mail$_{\text{Mass}}$ vs letter(s)'$_{\text{Count}}$ remain undiscussed here (Wierzbicka 1985; Middleton et al. 2004; Wisniewski 2010: 168; see Levin & Rappaport Hovav 2011 on such mass–count doublets). Instead this section aims to give a brief overview of the way mass nouns are distinguished from other nominal subcategories across languages, using as the key diagnostic feature the compulsory employment of a measure construction when a numerically unmarked noun is modified by a cardinal numeral (see Greenberg 1972: 16). The requirement that the noun must occur in its unmarked form excludes plural count nouns like 'books' or 'flowers' in expressions like *two boxes of books* or *two bunches of flowers*, where 'box' and the collective noun 'bunch' might also be interpreted as measure terms. The same requirement also means that pluralized mass nouns like 'wines' are excluded in, for example, *We tasted three or four wines*, because here 'wines' refers to different, countable types of the mass entity (see Dik (1997: 14) on 'subcategorial conversion'; see also Doetjes 2012 and 2017). Languages that are deemed to lack a separate category of mass nouns are discussed in sections 17.4.2 and 17.4.3.

C17P81 The actual form of the measure construction can vary from language to language. For example, whereas English uses a measure term with a preposition, like 'cup of' in *two cups of tea*, the preposition is absent in Dutch *twee koppen thee*, lit. 'two cups tea'. Mass nouns in classifier languages also require a measure construction, which in these languages involves the use of a mensural rather than a sortal classifier (Cheng & Sybesma 1998; Cheng 2012). Differences between measure terms and mensural classifiers are discussed in e.g. Allan 1977, Grinevald 2005, and Doetjes 2012.

Thai (Hundius & Kölver 1983: 170)

(20) *náamtaan* *săam* *thûaj*
sugar three cup
'three cups of sugar'

C17P82 Across languages there are various ways to express the portions of a linguistic mass entity (see Koptjevskaja-Tamm 2001 for an overview), in addition to (a) the construction that involves the use of an adposition as exemplified by the English *two cups of tea*, (b) the 'zero-strategy' exemplified by Dutch *twee koppen thee* (lit. 'two cups tea'), and (c) the variant that involves a mensural classifier (as in the Thai example (20)), there are also languages that use case marking, as in:

Russian (Slavic) (Koptjevskaja-Tamm 2001: 524)

(21) *čaška* *čaja/čaju*
cup:NOM tea:GEN/tea:PRTV
'a cup of tea'

C17P83 It has been argued that measure terms like Italian *manciata* 'fistful' in *una manciata di riso* ('a fistful of rice') or *bottiglia* 'bottle' in *una bottiglia di whisky* ('a bottle of whisky') belong to the category of so-called light nouns (Simone & Masini 2014), i.e. the nouns that are less than fully referential in certain syntactic contexts, in particular when they occur as part of a binominal construction as in the two examples above (on binominal NPs, see e.g. Foolen 2004 or Kim & Sells 2014). In such constructions, it is sometimes not clear which noun is the semantic head (Brems 2003; Doetjes 2017). Notice, finally, that classifiers have also been analysed as more or less grammaticalized elements, with sortal classifiers perhaps more at the more grammatical end of the continuum than mensural classifiers (Craig 1992).

C17S15 ### 17.3.4 The role of Shape and the place of nouns in an implicational hierarchy of word classes

C17P84 It was shown above that nouns with a positive or a negative value for the feature Shape appear to correlate with various grammatical properties. Here we briefly mention yet another area where the feature Shape appears to play a significant role, namely the parts-of-speech hierarchy proposed in Hengeveld (1992b: 68):

(22) Verb > Noun > Adjective > (manner) Adverb

C17P85 This hierarchy states, for example, that languages with a distinct class of adjectives must also have distinct classes of nouns and verbs and, importantly, that it may or may not have a dedicated class of manner adverbs. Thus, the implication only works one way: having a distinct class of nouns implies the language also has a distinct class of verbs, but it does not necessarily mean that the same language also has distinct classes of adjectives and manner adverbs.

C17P86 An important question is: what determines whether a language can have the next word class in the hierarchy shown in (22)? Restricting ourselves momentarily to nouns and adjectives, data from a representative sample of the world's languages indicate that all languages with a distinct class of adjectives also employ nouns which can be counted directly, i.e. +Shape nouns (Figure 17.1). In other words, it seems that the implication 'if a language has adjectives, then it has nouns', can now be reformulated more precisely as follows: 'if a language has adjectives, then it has +Shape nouns' (Rijkhoff 1999). The reverse is not true, i.e. languages with + Shape nouns do not necessarily have a distinct class of adjectives. For example, both Dutch and Lango have + Shape nouns (examples (6)–(8)), but whereas Dutch has distinct classes of verbs, nouns, and adjectives, Lango only has distinct classes of verbs and nouns and lacks a dedicated class of adjectives (Noonan 1992: 103). Speakers of Thai use −Shape nouns when they talk about objects like umbrellas and lemons, therefore nouns in this language lack the necessary positive value for the feature Shape (+Shape) that would allow for the next lexical category in the hierarchy: a distinct class of adjectives.

C17P87 Cross-linguistic data suggest that the role of the feature Shape is part of a larger picture, involving the coding of prototypical features of entities in the lexical meaning of verbs, nouns, and adjectives (Rijkhoff 2003 and 2008). Apparently, languages can only have these major word classes, if the basic meaning of the content words ~~encode~~ prototypical properties

of temporal ~~and~~ spatial entities (events and things). A prototypical event is a transitive activity involving an agent and a patient; a prototypical thing is a concrete, physical object. Thus, a language can only have major, distinct classes of verbs, nouns and adjectives, if the lexicon contains (a) lexemes that designate a dynamic relationship between an agent and a patient (i.e. transitive verbs), and (b) lexemes that designate a property that is specified as having a boundary in the spatial dimension (i.e. +Shape nouns). A modified version of the hierarchy that includes the necessary (but not: sufficient) semantic features looks as follows (see Rijkhoff 1999 and 2003 for details, with examples of languages occupying different positions in the hierarchy):

(23)	Verb	> Noun	> Adjective	>manner Adverb
	CONDITION:	$V_{+Transitive}$	$\mathbf{N_{+Shape}}$	$?A_{+Gradable}$

C17P88 The question mark under manner Adverb indicates that more cross-linguistic research is needed to determine if it is the feature + Gradable in the lexical meaning of adjectives that is required to allow for the occurrence of next distinct word class: manner Adverbs (see Wierzbicka 1986: 374–377; Beck 2011; Constantinescu 2015).

C17S16

17.4 Other part-of-speech systems and language change

C17P89 In addition to the nouns that belong to the six nominal subcategories discussed in sections 17.3.2 and 17.3.3, there are lexemes denoting properties of spatial entities that do not quite fit the six-way classification presented in Figure 17.1. On closer inspection, such lexemes belong to at least three different groups:

C17P90 (a) true nouns that are in the process of changing nominal subcategory (section 17.4.1);

C17P91 (b) words which, properly speaking, are not dedicated nouns but lexemes that belong to a flexible word class (section 17.4.2).

C17P92 (c) true nouns that are underspecified for the count vs mass distinction (section 17.4.3). Since these 'count–mass' nouns can be characterized in terms of Shape and Homogeneity, they could be added to the classification of nominal subcategories in Figure 17.1.

C17P93 Interestingly, measure constructions (section ~~17.3.3~~) seem to be absent in languages with lexemes of types (b) and (c); on this topic see, e.g. Greenberg 1972: 17).

C17S17

17.4.1 Nominal subcategories and language change

C17P94 The role of language change can be illustrated with examples from Mandarin Chinese and Hmong Njua (see also n. 8 on gradience and continuity between linguistic categories). Both

languages have what some have called plural markers and thus seem to contradict the idea that numeral classifiers (which appear with sort nouns) and plural markers (which appear on singular object and collective nouns) are mutually exclusive (see Xu 2012 for more discussion on this topic). It was already proposed in section 17.3.3.1 that human nouns in Mandarin are changing (or have changed) from sort noun to set noun and that the optional 'plural' suffix *-men* is better analysed as a collective aspect marker. Similar things can be said about Hmong Njua (Rijkhoff 1999: 241–242), which has sortal classifiers and where the collective classifier *cov* is said to be in the process of replacing all other classifiers to express 'plurality' (Harriehausen 1990: 115–117). According to Ratliff (1991), however, *cov* has a collective meaning and most probably derives from *cɔy*^{51} 'bunches or clusters of fruit' (superscript 51 indicates tone marks). In this case, too, one might argue that the nouns in question are changing membership from transnumeral sort noun to transnumeral set noun. Ultimately, these set nouns may turn into singular object nouns, as elements marking collectivity are a well-known diachronic source of plural markers (section 17.3.2.2; Rijkhoff 2004: 116–117).

C17S18

17.4.2 The absence of the so-called count–mass distinction in languages with 'nominals' (a flexible word class of Australian languages)

C17P95 In quite a few languages, problems in assigning lexemes to one of the six nominal subtypes in Figure 17.1 may be attributed to the fact that these languages lack a distinct class of nouns (recall that this chapter is only concerned with languages in which nouns constitute a separate word class), but instead have lexemes that belong to a so-called flexible word class (Rijkhoff & van Lier 2013). Languages with a flexible word class not only include languages with a single class of lexemes (called 'contentives' in Hengeveld 1992b, 2013) like Samoan or Tongan (see section 17.1), in which there is no clear distinction between verbs, nouns and members of other lexical categories, but also languages with a word class that Hengeveld coined 'non-verbs'. These languages have two major lexical word classes: (a) a distinct class of verbs and (b) a flexible class of non-verbs whose members translate in other languages as a noun or as an adjective (or even as a manner adverb). Such flexible lexemes are attested, for example, in Australian languages, many of which have a word class called 'nominals', which translate as nouns or adjectives in English (Dixon 1980: 271; Donaldson 1980: 70–71; McGregor 2013: 222; McGregor 2019b). The evidence suggests that members of the flexible class of non-verbs are not lexically coded to specifically denote a mass or a collective. Perhaps this should not be surprising: if a language has no dedicated class of nouns, nominal subcategories can at best be a marginal phenomenon.

C17S19

17.4.3 The absence of a count–mass distinction in languages in languages with a separate class of nouns

C17P96 There are quite a few languages in which notional count and mass nouns essentially share the same 'mini-grammar' (Wiltschko 2012; Lima 2018). Whereas members of the Australian

class of 'nominals' are strictly speaking flexible lexemes rather than dedicated nouns (section 17.4.2), many native North American languages have count–mass words that appear to be true nouns. These languages also lack a separate class of adjectives (just like many Australian languages), but here adjectival notions ('qualities') are expressed through verbs or nouns (Mithun 1999: 56). Another conspicuous difference between 'non-verbs' or 'nominals' in Australian languages and count–mass nouns in native North American languages concerns plural marking. In some North American Indian languages with a category of count–mass nouns, all members can appear with a plural marker. For example, in Ojibwe, where the count vs mass distinction is absent, all nouns can be pluralized (Rhodes 1990: 153–154; as reported in Corbett 2000: 87; for similar cases, see e.g. Mithun (1999: 80)).

C17P97 In Ojibwa, there is no grammatical distinction like the mass/count distinction of Indo-European. Thus, *mkwam* can equally mean 'ice' or 'a piece of ice'. *Nbiish* can mean 'water' or 'an amount of water'.

C17P98 Nouns which might be expected not to have a plural do in fact form plurals freely, interestingly with the unit reading and not with the sort reading. Thus *mkwam* 'ice' or 'piece of ice', *mkwamiig* (plural) 'pieces of ice'. Rhodes is unable to find a noun that cannot be pluralized in Ojibway.

C17P99 The absence of a count–mass distinction in nouns was also noted in Greenberg (1972: 16–17, referring to Whorf 1941: 80), who wrote that Hopi and many other native American Indian languages have no formally distinct category of mass nouns, as shown by the fact that, for example, numerals can directly modify nouns used to talk about masses and nouns used for countable objects: 'One says [in Hopi] not "a glass of water" but *kə̀ yí* "a water" […] not "a piece of meat" but *sikwí* "a meat"'. The fact that it is possible to say something like 'I brought them two water' (i.e. without a mensural construction) in languages with a category of count–mass nouns could be accounted for by assuming that the cultural or physical context would give sufficient clues about the appropriate container for the occasion. This is explicitly stated to be the case in, for example, the Brazilian language Yudja.

Yudja (Tupi) (Lima 2018: 7, 13–14)

(24) *Una* *txabïu* *apeta* *izaku*
1S three blood see
'I saw three units of blood' (the unit—drops, puddles, or containers—is identified by the context; in this case the speaker was referring to drops)

(25) *Una* *txabïu* *awïla* *wãẽ* *he* *izaku*
1S three honey pan in see
'I saw three pans of honey' or 'I saw portions of honey in a pan' (with a container like 'pan' the sentence may interpret both as referring to the units of counting (e.g. 'I saw three pans that contained honey') or as the location of these units)

C17P100 This is probably not very different from wat happens when a speaker of Dutch orders coffee in a café or restaurant, as in: '*Mag ik één/twee/drie/… koffie*?' (lit. 'May I one/two/three/… coffee?') 'I'd like to order one/two/three/… cups of coffee' where the Dutch mass noun *koffie* 'coffee' is directly modified by a numeral, i.e. without the usual measure construction, which requires a measure term like 'cup' as in *two cups of coffee*. Such a request

is perfectly acceptable in restricted contexts ('pragmatic coercion') and it is clear that the unexpressed measure here is 'cup' (Greenberg 1972: 16; Lyons 1977: 463; see LaPolla 2020 for an overview on the role of context in linguistics in the last two centuries; also Malinowski 1923).

C17P101 As mentioned earlier, one could assume that in languages like Ojibwa the count–mass lexemes constitute a underspecified nominal subcategory in their own right with neutral values for the features Shape and Homogeneity, which would suggest that these count–mass nouns are simply characterized as designating properties of spatial entities, including spatial entities with or without a perceptual distinct spatial contour (±Shape) as well as entities with or without an internal structure (±Homogeneity).

C17S20

17.5. Other properties of languages with a dedicated class of nouns

C17P102 The previous sections offered a predominantly lexically meaning-based classification of nouns with some attention for certain associated formal properties (such as the employment of number markers, nominal aspect markers or numeral classifiers) to the extent that they reflect semantic features of the various nominal subcategories. This section briefly mentions some other general properties of nouns, and how some of these properties are interconnected, beginning with genders and noun classes.

C17P103 Systems of nominal categorization occupy an important place in the literature on nouns, as only true nouns can be assigned to a grammatical gender or noun class (Rijkhoff 2008: 730). If a language has genders or some other system of noun categorization, it must have a distinct class of nouns. However, the reverse does not hold, as there are many languages with nouns but without nominal genders or noun classes, including English. Another important generalization concerning gender or noun class was proposed by Greenberg (1966: 95): '*Universal 36*. If a language has a category of gender, it always has the category of number'. Nouns can also be classified in terms of parameters like Animacy, Incorporability (Woodbury 1975) or Possession (e.g. alienably vs inalienably possessed nouns; Chappell & McGregor 1996). Due to space limitations, such cases are not discussed here; instead the reader is referred to the large number of detailed publications on gender and other systems of nominal classification.[14]

C17P104 Greenberg (1966: 93) also put forward a universal about the ordering of derivation and inflectional affixes on nouns and other lexemes. Universal 29 reads as follows: 'If both the derivation and inflection follow the root, or both precede the root, the derivation is always between the root and the inflection'. Morphological and other properties of the noun are also discussed in Lehmann & Moravcsik (2000); general, cross-linguistic overviews of the way new (derived) nouns are formed (including locational nouns) can be found in Bauer (2007, 2013) and Lieber (2017), in some languages, nouns are classified on the basis

[14] See, for example, Fodor 1959; Greenberg 1978; Craig 1986; Corbett 1991; Aikhenvald 2000 and 2006; Unterbeck et al. 2000; Seifart 2010; Corbett 2014; McGregor & Wichmann 2018. A comprehensive bibliography on this topic by Alexandra Aikhenvald (2011) can be found under the heading 'Classifiers' in *Oxford Bibliographies Online* (http://www.oxfordbibliographies.com).

of morphophonological criteria that determine the form of the appropriate plural marker (Rijkhoff 2004: 79–80; Terrill 2003: 105, 112–125). Furthermore, in Vinmavis and some other Oceanic languages 'the majority of nouns have initial *n*-, which is historically not part of the root' (Crowley 1985: 165). This is due to grammaticalization of an erstwhile determiner that gradually became an integral part of the noun (on this topic, see also Greenberg 1981 and 1991; Hoskison 1983: 24). It also appears that nouns and verbs can display distinct phonological behaviour. More specifically, in certain languages nouns appear to show more phonological contrasts than verbs. For example, Spanish nouns, but not verbs, have contrastive stress location (Smith 1997, 2011).

C17P105 There has also been a considerable amount of neuroscientific research on the way verbs and nouns are processed in the brain; for a recent overview see Vigliocco et al. 2011 (also Chapter 44 in this ~~volume), in~~ the area of first language acquisition, it has been argued that there is a (possibly language specific) noun bias, in that children seem to produce more nouns than verbs in the early stages of first language acquisition (see Childers 2014). Another difference between nouns and verbs is reported in a study by Seifart et al. (2012), who found that 'there is a robust cross-linguistic tendency for slower speech before nouns compared with verbs, both in terms of slower articulation and more pauses'. They argue that this slowdown effect is due to the increased amount of planning that nouns require compared with verbs.

C17S21

17.6 Conclusion

C17P106 This chapter has offered a cross-linguistic classification of basic (underived, unmarked) nouns that are used to talk about spatial objects, and which is based on the lexical features Shape and Homogeneity (section 17.3.1). It may be good to emphasize once again that this concerns a linguistic classification of nouns and not a classification of physical objects in the real world. Four nominal subcategories are characterized by a combination of positive or negative values for the lexical features Shape and Homogeneity ('Likepartedness'): singular object nouns (+Shape, −Homogeneity), collective nouns (+Shape, +Homogeneity), sort nouns, (−Shape, −Homogeneity), and mass nouns (−Shape, +Homogeneity). Nouns characterized by the feature + Shape designate a property with a spatial outline, which means they can be counted directly; nouns that have a negative value for the feature Shape (−Shape) require an extra, individualizing element (e.g. a sortal or mensural classifier) when they are modified by a cardinal numeral. Two nominal subcategories in this classification have a neutral value the feature Homogeneity: set nouns (+Shape) and general nouns (−Shape). Some languages appear to have an undifferentiated nominal subcategory, whose members can be used to talk about objects as well as masses (section 17.4.3). This ~~seem~~ to indicate that these mass–count nouns have neutral values for both Shape and Homogeneity.