Part 1: **Form (morphosyntactic strategies)**

Steve Pepper

Defining and typologizing binominal lexemes

Abstract: This chapter starts by demonstrating the need for the comparative concept 'binominal lexeme' in order to cover both 'noun-noun compounds' and their 'functional equivalents' (§1). To complement this informal definition, four different, but compatible definitions of binominal lexeme are developed: functional, onomasiological, formal and typological (§2). Although couched in a variety of terms based on different theoretical frameworks, these have essentially identical extensions.

In §3 a nine-way classification of binominal strategies is presented, together with the mnemonics used throughout this volume: **jxt**, **cmp**, **der**, **cls**; **prp**, **gen**, **adj**, **con**, and **dbl**. These nine types are represented on a two-dimensional grid that captures the number of markers, the locus of marking and the degree of fusion. The grid reveals two lacunae or "missing types": **prn** and **nml**. Whereas the first of these probably exists somewhere in the world's languages, the second seems to be a logical impossibility.

§4 discusses types that are intermediate between the nine main types and the grammaticalization pathways that produce them. It then goes on to examine the relationship between binominal constructions and adnominal possessives, and introduces a new methodology, based on the Pwav scale, for comparing two non-binary constructions. This leads to the formulation of two Greenbergian universals concerning binominals and nominal modification.[1]

1 Introduction

'Word-formation' – one of the two branches of 'morphology' (the other being 'inflection') – has traditionally been subdivided into 'compounding' and 'derivation'. Recent research, however, has shown these distinctions to be grossly over-simplified and misleading (Bauer 2005), to the point that many important linguistic phenomena that fall between the two stools tend to be overlooked. This chapter provides an overview of one such phenomenon: the process of forming new lexemes by combining two (or more) existing lexemes that denote nominal concepts.

1 For language names, ISO codes and genealogical classifications, as well as sources for all language examples, see Pepper (2020).

Note: This chapter has been made Open Access *in memoriam* my parents Harry Pepper (1926–1996) and Edna Pepper (1932–2022).

https://doi.org/10.1515/9783110673494-002

The most well-studied strategy for this kind of lexeme formation is the kind of noun-noun compounding found in Germanic languages (1).

(1) German (DEU) *Eisenbahn* [iron.way] 'railway'

Crosslinguistically, however, there are many other strategies, including, among much else, the use of a **preposition** (2a), an **adjectivizer** (2b) a **dependent-marking affix** (2c), and a **head-marking affix** (2d).

(2) a. French (FRA) *chemin de fer* [way PREP iron]
b. Russian (RUS) *železnaja doroga* [iron.ADJZ road]
c. Bezhta (KAP) *kilos hino* [iron.OBL.GEN way]
d. Turkish (TUR) *demir yolu* [iron road.3SG]

Like *Eisenbahn*, all the examples in (2) combine the concepts IRON and WAY in order to denote the meaning 'railway', but they do so using quite different morphosyntactic strategies. However, since they are all "phrasal" in nature, they fall outside the domain of 'morphology', and consequently also 'word-formation', as traditionally understood. And since they are lexical, they tend to fall outside the domain of 'syntax' as well. They fall between two stools.

To cite a few examples from the otherwise excellent five-volume handbook of word-formation in the languages of Europe (Müller et al. 2015): Floricic (2016) limits his coverage of the prepositional type (*chemin de fer*) – by far the most common way to form new lexemes in French – to a single sentence. Uluhanov (2016) makes no mention at all of the adjectival type (*železnaja doroga*) in Russian (relational adjectives are only mentioned in the context of denominal derivation). Nor do Khalilov and Khalilova (2016) mention the genitival type (*kilos hino*) in Bezhta.

In contrast, Wilkens (2016: 3370) treats the Turkish type (*demir yolu)* under "composition" (i.e., compounding), showing how *ev kapısı* [house door.3SG] 'front-door' "displays one basic feature of Turkish composition with the compound marker *-(s)I* (3rd person singular possessive suffix)".

This uneven treatment is the result of two unhappy circumstances: the division of grammar into 'syntax' and 'morphology', and the use of language-specific categories for the purpose of cross-linguistic comparison. What is clearly a compound in German (1), and equally clearly a compound (according to the local tradition) in Turkish (2d), is marginal at best in French (2a), of unclear status in Bezhta (2c) and can in no way be considered a compound in Russian (2b). And yet all five forms are functionally and semantically equivalent: they all consist basically of two nouns – one of which modifies the other, which denote the concepts IRON and WAY respectively, and they all denote the same "complex concept" (see

Masini, Mattiola & Pepper, this volume): RAILWAY. They differ only in the morphosyntactic strategy employed to combine the two nominal roots.

It is in order to bring these five strategies (along with others) under a single umbrella that the comparative concept 'binominal lexeme' was developed (Pepper 2020). Informally, binominal lexemes (or 'binominals' for short) are *noun-noun compounds or their functional equivalents*.

The goal of this chapter is to provide a general introduction to binominals. Section 2 offers four different definitions of binominal lexeme. Although couched in a variety of terms, based on different theoretical frameworks, these have essentially identical extensions. Section 3 then presents a nine-way classification of binominals based on the morphosyntactic strategies that they employ in the world's languages. Five of these were exemplified in (1) and (2), and four more will be introduced in §3.2. These nine 'binominal types' are arranged on a novel two-dimensional grid that also captures the number of grammatical markers, the locus of marking, and the degree of fusion. In addition, the grid reveals two theoretically possible, but unattested strategies, which are discussed in §3.3. Finally, section 3 discusses binominals in the context of grammaticalization, both in terms of gradient binominal phenomena (§4.1) and the relationship between binominals and adnominal possessives, or – more generally – nominal modification constructions (§4.2). In order to do so, it describes an innovate general method for comparing two non-binary constructions.

2 Defining binominal lexemes

While fairly accurate, the preliminary definition of binominal lexeme given above – as a noun-noun compound or its functional equivalent – is too imprecise for cross-linguistic comparison. In this section I offer more precise definitions that can be broadly characterized as functional (§2.1), onomasiological (§2.2), formal (§2.3) and typological (§2.4).

2.1 The functional definition

We may start by asking, what in fact *is* the function of a noun-noun compound? The answer, it would seem, is to name a new concept through combined reference to two existing nominal concepts. This is a 'comparative concept' (Haspelmath 2010) that is suitable for cross-linguistic comparison, since it is based solely on functional and semantic, rather than formal, criteria.

The naming function is important, since this is what distinguishes a binominal, such as *chemin de fer*, from an adnominal possessive, or 'adpossessive', such as Fr. *la plume de ma tante* 'the pen of my aunt'.

Specifying that the two combining concepts are nominal serves to exclude adjective-noun combinations like Eng. *blackbird*, and verb-noun combinations like It. *lavapiatti* 'dishwasher' [lit. 'wash dishes'], which are clearly not noun-noun compounds. The term 'nominal' is used here in the extralinguistic sense of pertaining to an object, as opposed to an action or a property.) However, it has further significance, in that it excludes forms like Eng. *walking stick* and *dishwasher* – i.e. so-called synthetic compounds, which are usually regarded as noun-noun compounds. In each of these, one of the combining concepts (represented by the modifying and modified elements, respectively) refers to an action (walking, washing) rather than an object.

The reason for excluding synthetic compounds and the like from the category 'binominal lexeme', is the suspicion that they may exhibit deviant behaviour, and that their inclusion could potentially "muddy the waters" of the analysis. One way in which they clearly *do* differ from binominals is the following. In binominals, the relationship between the two combining concepts is unstated (or at least underspecified): the motivation for combining the concepts WAY and IRON in *chemin de fer* is because a railway is regarded as a 'way' that is *made (*or *composed) of* iron. However, that relationship is not stated explicitly. In *dishwasher* (and *lavapiatti*), on the other hand, the relationship between the instrument (denoted by the suffix *-er*) and the dishes is explicit: it is a washing relation.

This leads us to the first, purely *functional* definition of binominal lexeme:[2]

(3) **A binominal lexeme is a naming unit that is based on two nominal concepts one of which modifies the other.**

This functional definition can be expressed more formally in three different ways.

2.2 The onomasiological definition

We start with the onomasiological definition, because Štekauer's (1998) system of 'onomasiological types' (OTs) was seminal to the development of the comparative

2 The rationale for including the qualification "one of which modifies the other" is explained below.

concept of binominal lexeme (see Pepper 2020: §1.2.3).[3] Štekauer's classification is largely unfamiliar to mainstream linguistics (less so in Europe than elsewhere), primarily because it employs "non-standard" terminology that has its roots in the Prague School of Linguistics, in particular, the work of Miloš Dokulil (1962; 1966; 1994), most of which is available only in Czech (CES).[4] This lack of familiarity is to be regretted, since Štekauer's work contains many interesting insights of interest to morphologists and typologists. To help rectify this unfortunate state of affairs, this section will attempt to "translate" Štekauer's theory into more familiar terms.

The onomasiological typology is based on the recognition of four 'conceptual categories': SUBSTANCE, ACTION, QUALITY, and CONCOMITANT CIRCUMSTANCE (Štekauer 1998: 9). The first three of these are directly equivalent to Croft's semantic classes OBJECT, ACTION and PROPERTY (Croft 2022, cf. also many earlier works); the latter terms are more familiar and will therefore be used here in preference to Štekauer's.[5] Furthermore, Štekauer's exposition pertains to the linguistic sign in general, of which complex nominals are but a subtype; in line with the focus of the present volume, and since most of Štekauer's examples are in fact complex nominals, the present discussion is restricted to the latter.

The semantic structure of a complex nominal consists, in principle, of two parts: "an *onomasiological base* denoting a class, gender, species, etc., to which the object belongs", and "an *onomasiological mark* which specifies the base" (Štekauer 1998: 9). These terms translate directly into 'head' and 'modifier', respectively. Both of these elements represent one of the above-mentioned conceptual categories. In the case of complex nominals, the head almost always represents an object (or 'substance').

The modifier can be either simple or complex. A simple modifier represents a property, e.g. *black* in *blackbird*. A fully specified complex modifier consists of a verbal ('actional') part and an argument to the verb; following Dokulil, Štekauer calls these the 'determined' and the 'determining' constituents. In *man-eating tiger* the modifier *man-eating* consists of a determined constituent *eat* and the determining constituent *man* (that which is eaten). These equate to Croft's (2022) 'property modification construction' and 'action modification construction',

3 Štekauer's classification is embedded in a broader theory of word-formation that is interesting in its own right but not directly relevant to the present discussion.

4 Note, however that the 1962 work contains a 31-page summary in English, and that the 1966 and 1994 papers are in German and English, respectively.

5 Langacker (e.g. 2008), Haspelmath (e.g. 2012) and many others use the term 'thing' instead of 'object'. CONCOMITANT CIRCUMSTANCE covers Place, Time, Manner, etc. See §2.4 for more details on Croft's model.

respectively (see §2.4), and they constitute two of Štekauer's onomasiological types (see Table 1).

Now, in the act of word-formation, when this semantic structure is given linguistic form, part of a complex modifier, either the determined (actional) part or the determining part, may be omitted. In *spinning wheel* there is no determining element, and in *summer house* there is no determined (actional) element. These constitute two more onomasiological types.

Table 1: Onomasiological types 1–4.

Onomasiological type (OT)	Example	Modifier			Head
		Simple	Complex		
			Determined	Determiner	
OT1	*man-eating tiger*	n/a	EAT	MAN	TIGER
OT2	*spinning wheel*	n/a	SPIN	–	WHEEL
OT3	*summer house*	n/a	–	SUMMER	HOUSE
OT4	*blackbird*	BLACK	n/a	n/a	BIRD

There is a fifth onomasiological type, OT5, which is characterized by "an unstructured onomasiological level" (Štekauer 2005: 221) and covers cases of conversion, such as $time_N$ ~ $time_V$. This type is not relevant to the present discussion.[6]

Based on the preceding, it is clear that binominals correspond to Onomasiological Type 3. In Štekauer's terms, our original example (1), *Eisenbahn*, is a naming unit that consists of the onomasiological base (*Bahn*) and the determining element of the (complex) onomasiological mark (*Eisen*). Expressed in more familiar terms, *Eisenbahn* is a complex nominal that consists of a head (*Bahn*) and a (conceptually complex) modifier consisting of a determining element (*Eisen*); the determined element, representing the semantic relation between the two, is unstated.

We can thus formulate the following *onomasiological* definition of binominal lexeme:

(4) **A binominal lexeme is an Onomasiological Type 3 naming unit.**

The adoption of an onomasiological perspective has important implications for the interpretation of the functional definition (3). This is because the onomasiological

6 For a more in-depth presentation, critique and revision of the system of onomasiological types, see Pepper (2018; in prep.).

model accords the same status to derivational affixes and lexical roots. In theory, an affix can represent any of the elements of the onomasiological structure:

- In *house-keeping*, the nominalizing suffix *-ing* denotes a process and represents the onomasiological base (i.e., the head), *keep* is the determined element of the mark (i.e., the modifier), and is *house* the determining element of the mark; so this form is OT1.
- In *writer*, the agentive suffix *-er* represents the base and *write* the determined element of the mark; so this is OT2.
- And in *novelist*, the agentive suffix *-ist* represents the base and *novel* the determining element of the mark; so this is OT3 – and consequently also a binominal.

Now, so far in this chapter, every example of a binominal lexeme has consisted of two nouns (in addition, sometimes, to additional grammatical material, such as the preposition *de* in *chemin de fer*). The consequence of adopting the onomasiological perspective is that a binominal lexeme may be comprised not only of two nouns, but of a noun and a nominalizer.

Furthermore, a binominal may consist of a combination of a noun and a noun classifier or noun class marker, as in Bora (BOA) *túúheju* [nose.CLF:hole] 'nostril' (Urban 2012: 127), Harakmbut (AMR) *siro-pi* [metal-CLF:stick] 'knife' (Rose and Van linden, this volume) or Bandial (BQJ) *jijamen* [CL:ji.goat] 'kid' (Watson 2015). However, this applies only when the classifier or class marker in question has a derivational function; if the function is merely classificatory, as in Bandial *ejamen* [CL:e.goat] 'goat', the form is not considered to be a binominal lexeme (see the discussion of the **cls** type in §3.2.1.4 below and Rose and Van linden, this volume).

2.3 The formal definition

It is possible to develop a further definition of binominal lexeme, one that may be more accessible for some linguists, in the following way. If we (provisionally) ignore the refinement of the notion of binominal to include denominal derivations (e.g. *novelist*) and certain noun classifier constructions (e.g. *siro-pi*), a simple definition of binominal would be "a naming unit consisting of two nouns, and possible additional grammatical material". However, in addition to being incomplete, this definition suffers from the problem that 'noun' is not a well-defined cross-linguistic comparative concept (Haspelmath 2012). The latter issue can be addressed using Haspelmath's term 'thing-root' (defined as "a root that denotes a physical object (animate or inanimate)") instead of 'noun', but this would still not encompass forms like *novelist* and *siro-pi*, since neither *-ist* nor *-pi* are roots. In

addition to the term 'thing-root', we therefore require the notion of 'thing-affix', defined as "an affix that denotes a physical object (animate or inanimate)"; this would cover both nominalizers like *-ist* and classifiers like *-pi*.[7]

Since roots and affixes are both morphs (Haspelmath 2020), a suitable cover term for thing-root and thing-affix is 'thing-morph', defined as "a morph that denotes a thing (prototypically a physical object, animate or inanimate)".[8]

Now, a binominal lexeme by its very nature involves an unstated (or underspecified) relation R between the two nominal concepts: the MADE OF relation in *chemin de fer* and *siro-pi*; the PART OF relation in *túúheju*; the CREATOR OF relation in *novelist*; etc. (see Pepper, this volume, b for further discussion). This semantic relation constitutes the motivation for combining the two concepts in question: just as a railway is conceptualized as a way that is made of iron, a knife is a stick made of iron, a nostril is a hole that is part of a nose, and a novelist is someone who writes (or more generally, ceates) novels.

This aspect of binominals can be usefully incorporated into its definition (5):

(5) **A binominal lexeme is a naming unit that consists primarily of two thing-morphs, and possibly additional grammatical material, formed by combining two concepts between which there is an unstated (or underspecified) relation of modification.**

In a sense the reference to the unstated relation is redundant, since every naming unit consisting of two thing-morphs involves such a relation. However, it serves to exclude so-called 'co-compounds' (Wälchli 2005), in which the relation between the two constituents is one of coordination rather than modification. In addition, it highlights the existence of the semantic relation, and it explicitly excludes forms such as Viet. *bữa ăn sáng* [meal eat morning] 'breakfast', in which the additional material (over and above the thing-morphs *bữa* and *sáng*) denotes an action, EAT, making this an instance of Onomasiological Type 1 (see §2.2 above).

In conclusion, a binominal lexeme can take any of the following forms:

– two nouns (e.g. *rail.way*) – possibly with additional grammatical material (e.g. *chemin de fer*);
– a noun and a nominalizing affix (e.g. *novel.ist*) – possibly with additional grammatical material (e.g. Slovak (SLK) *železn.ica* [iron.ADJZ.NMLZ] 'railway');

7 This assumes, of course, that the definition of affix covers classifiers and class markers, but we leave that issue aside here.

8 The addition of the qualifying "prototypically" allows for the extension of Haspelmath's concepts to also cover abstract 'objects', such as LOVE.

– a noun and a classifier (e.g. *siro-pi*) – possibly with additional grammatical material (e.g. Harakmbut *wã̲-õh-wẽ* [NPF-nose-CLF:liquid] 'nostril';
– arguably, two nominal affixes, as in neoclassical compounds, e.g. *hydro-mancy* < water + divination (see §3.2.1.3).

One might argue that this definition is unnecessarily restrictive, in that it excludes not just synthetic compounds and coordinate compounds, but also forms consisting of three (or more) thing-morphs denoting just two concepts (one complex concept, denoted by a binominal, and one simple concept, denoted by a thing-morph). Indonesian *jalan keréta api* 'railway' is a case in point, consisting as it does of *jalan* 'road' and *keréta api* [carriage fire] 'train'. However, the definition is more than sufficient for the purpose of the present volume.

2.4 The typological definition

As a comparative concept, the binominal lexeme construction can also be defined in terms of Croft's (1991: 67; 2001: 88; 1990: 185; 2022) model of basic cross-linguistic constructions. This model, the Scapa Grid,[9] is shown in Table 2. The model is based on Croft's insight that constructions can, indeed *must*, be defined cross-linguistically in terms of two parameters: semantics and "information packaging" (Croft 2022). In the Scapa Grid, these are realised as semantic classes and propositional acts, respectively.

Table 2: Croft's Scapa Grid of cross-linguistic constructions.

Semantic Class	**Propositional Act**		
	reference	**modification**	**predication**
object	UNMARKED NOUNS	genitive, adjectivizations, PPs on nouns	predicate nominals
property	deadjectival nouns	UNMARKED ADJECTIVES	predicate adjectives
action	action nominals, complements, infinitives, gerunds	participles, relative clauses	UNMARKED VERBS

9 Croft has used this table for over 30 years but has never given it a name. It is called here the "Scapa Grid", since the cells are at the intersection of three **S**emantic **C**lasses (object, property and action) **A**nd three **P**ropositional **A**cts (reference, modification and predication).

In terms of this model, binominal lexemes fit neatly into the cell **object** + **modification**, and since 'modification' for Croft always means modification of an object concept, this equates to what other linguists (e.g. Bauer & Tarasova 2013: 10) call "adnominal nominal modification". Croft (2022) himself adopts the term 'nominal modification construction' (see §4.2 below).

What distinguishes binominal lexeme constructions from other nominal modification constructions, such as the possessive (modification) construction, is that the former involve lexicalization (see §4.2). On this basis, the following *typological* definition can be stated:

(6) **A binominal lexeme is an instance of a lexicalized nominal modification construction.**

All of the preceding definitions – functional (3), onomasiological (4), formal (5), and typological (6) – have the same extension. They all include both denominal nominal derivations and they all exclude synthetic compounds. They also all exclude coordinate compounds: the onomasiological definition does so because Onomasiological Type 3 involves an onomasiological base (i.e., a head) and a determining element (i.e., a modifier), and the typological definition states clearly that the relation between the two elements is one of modification, which again implies a head and a modifier. The two other definitions, however, require the qualifications regarding the nature of the relation noted above.

Having now defined binominal lexemes in four different, but compatible ways, we can proceed to how they may be classified on the basis of the morphosyntactic strategies that they exhibit: the typology of binominal lexemes.

3 Classifying binominal lexemes

In this section we present the classification of binominals based on morphosyntactic strategies (Croft 2022) that was originally developed by Pepper (2020) and is used throughout the present volume.

3.1 Pepper's (2020) nine-way typology

Intuitively, binominal lexemes are closely related to possessive constructions, or more precisely, those that express adnominal possession – as opposed to predicative possession and external-possessive constructions (Koptjevskaja-Tamm

2002). Attention has already been drawn (in §2.1) to the role of French *de* 'of' in both the binominal *chemin de fer* and the possessive noun phrase (PNP) *la plume de ma tante*. The structure [A *de* B] is common to both French constructions, with A denoting the head or possessum and B denoting the modifier or possessor. Similarly, in Germanic languages the formative *-s-* can be either a binominal linking element or a genitive marker. (It is somewhat rare in Modern English binominals, but examples such as *women's magazine*, *ladies' man*, *dog's breakfast* and *wolf's bane* show that it does exist.)

In Russian binominals there is a strong tendency to incorporate an adjectivizer, as in *železnaja doroga* [iron.ADJZ road] 'railway', whereas most adnominal possessives utilize the genitive case. However, examples of Russian binominals that use the genitive do exist (e.g. *palec nogi* [digit foot.GEN] 'toe'), as do Russian adnominal possessives that utilize adjectivizers, as in *mojeho bratrowe dieci* [1SG:M.GEN brother.ADJZ.PL child.PL] 'my brother's children' (Corbett 1987: 300).

This sharing of the same morpheme in binominals and possessives is found in languages all across the world, as witnessed by the fact that Pepper's (2020) database of 3,738 binominals contains over 400 instances in which the gloss includes one of the abbreviations 3SG, POSS, AL and INAL (which by no means exhausts the list of possessive morphemes).

Other evidence for the close relationship between binominals and adnominal possessives includes a significant overlap between the kinds of semantic relation exhibited by the two (Pepper 2010; 2016), and the existence of a bidirectional word order universal **Poss-N** ≡ **Mod-N** between possessives and compounds in Bauer's (2001) data, drawn from a genetically and areally balanced sample of 36 languages (Pepper 2020: 27).

All of this evidence suggests the hypothesis that a binominal lexeme construction will often have grammaticalized out of an adnominal possessive construction. That being the case, it makes sense to base a typology of binominals on that of possessive constructions, since it will facilitate investigation of the grammaticalization hypothesis.

The most comprehensive typological work on adnominal possession is Koptjevskaja-Tamm's (2002; 2003; 2004) survey of possessive noun phrases (PNPs) in Europe, in which she develops the typological classification shown in Figure 1.

PNPs are here subdivided on the basis of fusion (synthetic vs. analytic), with juxtaposition in between. Synthetic PNPs are further subdivided by the locus of marking (on the head, the dependent, or both), and analytic PNPs are subdivided into those that employ prepositions and those that employ linking pronouns. The classification is actually more extensive than this, since mention is also made (2002: 144) of a seventh type, possessive compounding, which is "mainly

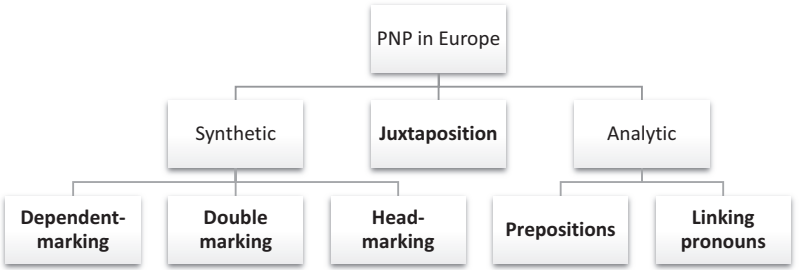

Figure 1: Koptjevskaja-Tamm's typology of ossessive noun phrases.

restricted to Northern Swedish". In addition, derived (relational) adjectives are mentioned in passing (2002: 157) but not included explicitly in the classification.[10]

Of Koptjevskaja-Tamm's eight types, seven are found in Pepper's (2020) binominals data. Linking pronouns are not found (but see below under *Unattested strategies*, §3.3.1). In addition, Pepper identifies two strategies for creating binominal lexemes that are not used for attributive possession: denominal derivation (e.g. *novelist*; cf. Pol. *wiatr.ak* [wind.NMLZ] 'windmill') and the derivational use of noun classifiers (e.g. *siro-pi*; *jijamen*).

In all, nine strategies were identified by Pepper. They are listed in Table 3, together with examples of each strategy, labels, and three-letter codes that will be used extensively, both in the following presentation (which discusses each type in detail) and throughout this volume.[11] Note that the three-letter codes and labels are simply mnemonics and should not be taken literally: they are intended to suggest the prototype of each category rather than its full generality. Thus, **prp**, for example, denotes a type in which the additional marker can be any independent lexeme that forms a constituent with the modifier (for example, a postposition, connector, linker or determiner) and not just a preposition. Similarly, the additional, non-transpositional (i.e. non-word-class changing) affix attached to the modifier in a **gen** strategy need not necessarily be a genitive marker (although this is the prototypical case).

In addition, types are grouped according to the degree of grammatical marking that they involve. Four strategies (**jxt**, **cmp**, **cls** and **der**) involve no additional linguistic material, over and above the two main constituents; another four (**adj**, **gen**,

10 Whether this is because they are also considered marginal, or because they can be subsumed under dependent-marking, is unclear.

11 Strategies marked with an asterisk (*) are mentioned by Koptjevskaja-Tamm but not included in her six-way classification; those marked with a dagger (†) are new.

Table 3: Nine binominal strategies (Pepper 2020).

Marking	Strategy	Code	Example
0	juxtaposition	**jxt**	VIE *đường sắt* [road iron] RAILWAY
	compounding *	**cmp**	DEU *eisen.bahn* [iron.way] RAILWAY
	classifier †	**cls**	BOA *túú.heju* [nose.CL(HOLE)] NOSTRIL
	derivational †	**der**	SLO *želez.nica* [iron.NMLZ] RAILWAY
1	adjectival *	**adj**	RUS *želez.naja doroga* [iron.ADJZ.NMLZ] RAILWAY
	genitival	**gen**	KAP *kil.os hino* [iron.GEN road] RAILWAY
	adpositional	**prp**	FRA *chemin de fer* [way PREP iron] RAILWAY
	construct	**con**	PLT *lala.m.by* [road.PER.iron] RAILWAY
2	double	**dbl**	TBC *-emo.lɨ sakɨla.lɨ* [nose.POSS aperture.POSS] NOSTRIL

prp and **con**) involve marking on either the head or the modifier; and one (**dbl**) involves marking on both the head and the modifier.

The original version of this typology, which was presented to contributors of the present volume at SLE 2016 in Zürich and is reflected in some of the contributions to the present volume, consisted of eight types instead of nine: juxtaposition (**jxt**) and compounding (**cmp**) were not distinguished, on the grounds that the presence or absence of a space between the two constituent nouns (or the use of a hyphen) is merely an orthographic convention – as witness the variable spelling in English of a binominal like *flowerpot ~ flower-pot ~ flower pot*, and the alternative ways of transliterating a Japanese binominal like 蜘蛛の巣 SPIDER WEB as either *kumo no su* [spider GEN web] or *kumonoso*.

However, this was inconsistent with the decision to recognise **prp** and **gen** as separate strategies, since these differ only in whether the additional relational marker is a separate word or an affix. Instead of merging these two strategies, which would obscure important facts in languages that exhibit both types, the **jxt** type was added to the typology – despite the well-known fact that there is no accepted cross-linguistic definition of the notion of word (Haspelmath 2011). This is justified by the fact that every language appears to have such a notion, or, as Bauer (2000: 255) puts it, "all languages have a unit which falls between the minimal sign and the phrase". As Koptjevskaja-Tamm (2004: 175) says:

> *Juxtaposition or compounding* The border between juxtaposition and compounding is notoriously difficult to draw, and much more research is needed for determining to what degree this distinction makes sense cross-linguistically. Until then in many cases we have to rely on the local tradition. Thus, Mordvin is traditionally described as resorting to juxtaposition for cases like *tuma lopa* 'an oak leaf' or *ved' vedra* 'a water pail', while their English correspondents are normally treated as compounds. Also, as well known, combinations of head nomi-

> nals with genitive-marked dependents and even with prepositional dependents, like Fr. *un chemin de fer* 'railway' (lit. 'a road of iron'), often border on compounds, and the absence of consensus on the treatment of cases like *women's magazine* and *boys' school* testifies to this.

In the absence of more robust criteria, we adopt Koptjevskaja-Tamm's policy of relying on the "local tradition". Given the nature of his data, Pepper (2020) employs the orthographic heuristic that a word space or hyphen signals juxtaposition (**jxt**), whereas the lack of either signals compounding (**cmp**). Pepper (2020: 257) found that this heuristic is actually sufficiently robust to reveal at least one interesting universal, viz. that two nouns are significantly more likely to fuse when the head is on the right.[12]

In addition, the original version of the typology did not stipulate how to handle cases in which multiple morphemes occur on one of the two main constituents. This is clarified by Pepper (2020: 142) as follows:[13]

> In order not to complicate the typology unnecessarily, two or more consecutive morphs attached to either the modifier or the head are counted as a single morph. For example Bezhta *kil.o.s hino* [iron.OBL.GEN way] RAILWAY is simplified to *kil.os hino* [iron.GEN way] and treated as having just one additional morph. In this I follow Nichols (1992: 62), who found such a simplification necessary "because the precise amount of multiple case marking in the constructions I am surveying is generally not made clear in grammars, so no consistent count could be made". A further reason is that there are too few examples of this phenomenon in my data to justify defining separate types to cater for them.

It is possible to arrange the resulting nine binominal types (or strategies) hierarchically, but to do so requires choosing between a grouping based on degree of fusion (i.e. analytical vs. synthetic, as in Figure 1), or one based on locus of marking (i.e. head, dependent, both or none), both of which would obscure important facts. Koptjevskaja-Tamm's decision to use the former obscures the fact that her category 'prepositional' is also a dependent-marking strategy, and the fact that markers on the head or dependent may be pronominal in nature. Similarly, a decision to use the latter would obscure other important facts.

In order to avoid such issues, the nine binominal types are arranged on a two-dimensional grid (Figure 2) which incorporates three different parameters: number of marked elements, locus of marking, and degree of fusion. It also allows

12 In Pepper's database, binominals of type **jxt** are evenly divided (in a ratio of 1:1) between right-headed and left-headed, whereas those of type **cmp** favour right-headedness by a ratio of more than 4:1.

13 This stipulation is particularly important in order to deal with polysynthetic languages like Anindilyakwa (see van Egmond, this volume).

for the addition of a second plane (or third dimension) if one wishes to capture the order of elements.

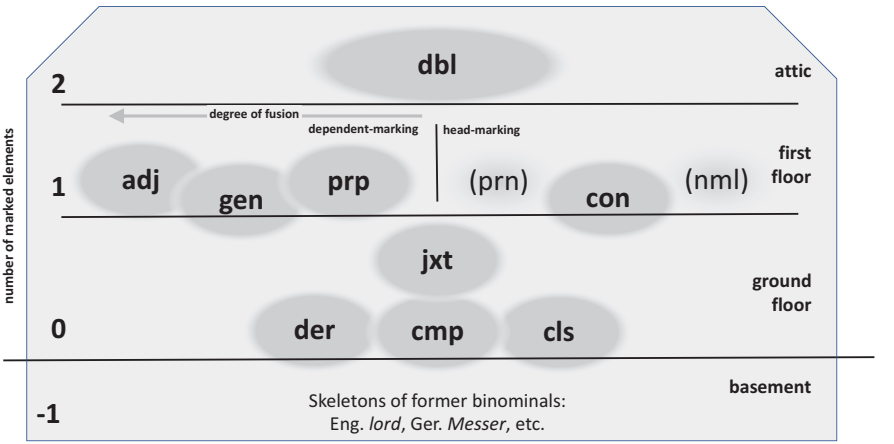

Figure 2: Visualization of the nine strategies on a grid.

The figure may be thought of as a two-storey house with an attic and a basement. On the ground floor (level 0) we find the four types that consist of just the two primary constituents and no additional grammatical material. On the first floor (level 1) are the four types which have an additional morpheme associated with one of the two primary constituents. (There are two apartments on this floor, one for head-marking and one for dependent-marking.) In the attic (level 2) there is just one type, with additional morphemes attached to each of the primary constituents. Finally, there is a basement containing the skeletons of former binominals such as Eng. *lord* and Ger. *Messer* (see §4.1.1).

The vertical dimension thus represents the *degree of marking* (0, 1 or 2 marked elements) and, on level 1, the horizontal dimension represents the *locus of marking* (head or dependent). In addition, the positioning of the three types in the left-hand apartment on level 1 (dependent-marking) represents the *degree of fusion*. The latter, illustrated in Figure 3, is a continuum that ranges, in Bybee's (1985: 12) words, from "the most highly fused means of expression, lexical expression", to "the most loosely joined means of expression, syntactic or periphrastic expression".

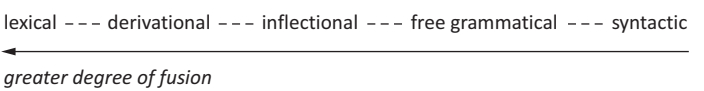

Figure 3: Degree of fusion (Bybee 1985).

The occupied room, **con**, in the right-hand apartment on the first floor (head-marking) is positioned in the middle in order to mirror **gen**, with which it shares the feature of involving a non-transpositional affix (as opposed to a transpositional affix or adposition).[14] Two of the rooms are vacant; since no binominals have been found representing types that mirror **adj** and **prp**, gaps are shown in those positions. These lacunae, labelled (**prn**) and (**nml**), represent the potential existence of two further types, as yet unattested:[15] an analytic form of head-marking (which would correspond to **prp**), and a type in which the head bears a transpositional affix (which would correspond to **adj**). These "missing types" are discussed below in §3.2.4.

The degree of fusion is also relevant to the vertical organisation on the ground floor, since compounds often evolve from juxtaposition and are in turn the source of classifier constructions and derivations. It might seem irrelevant to the organization of the attic, which only contains a single type. However, this is because **dbl** is something of a catch-all, which makes no distinctions in terms of function or degree of fusion. In theory, the two morphemes involved in a **dbl** construction could be a combination of any of the types found elsewhere in the model: freestanding or affixed, transpositional or non-transpositional, and a more fine-grained classification based on more data remains to be done.

The use of ovals with gradient fill rather than the boxes with sharp outlines found in more traditional representations (including Koptjevskaja-Tamm's) is intended to convey the fact that the types are not clearly defined categories but rather points within a multidimensional space at which phenomena tend to cluster. This enables the representation of overlaps, grammaticalization pathways and various "in-between" or gradient phenomena, as will be seen in §4.

One parameter that is not catered for in the basic typology is the order of elements or, as it is usually called in studies of compounding, the position of the head. Such a parameter is required to differentiate between left- and right-headed binominals of the same type within a particular language, such as the existence of both Head-Mod and Mod-Head **jxt** binominals in Vietnamese (VIE) (7a); and of both Head-Mod.ADJZ and Mod.ADJZ-Head **adj** binominals in Polish (POL) (7b). This parameter is orthogonal to those shown in Figure 2 and can be thought of as an additional plane that mirrors the one depicted in Figure 2. The additional

14 The parallel between these two types is underscored by the fact that the term 'genitive' is sometimes applied to possessive markers that attach to the head, as in Malagasy *lalandrà* [road.GEN.blood] 'vein or artery' (Adelaar 2009). Dixon (2010: 268) advocates the use of the term 'pertensive' for possessive markers on the possessum, but we follow Creissels (2017; this volume) in adopting the term 'construct (case)'.

15 But see van Egmond (this volume) and §3.3.2.

parameter extends the nine-way classification into an 18-way classification, with labels such as **jxtL**, **jxtR**, **adjL**, **adjR**, etc.

(7) a. Vietnamese
xe lửa [vehicle fire] TRAIN – **Head Mod** (native form)
hoả xa [fire vehicle] TRAIN – **Mod Head** (loan from Sinitic)
b. Polish
kolej żelaz.na [course iron.ADJZ] RAILWAY (arch.) – **Head Mod.ADJZ**
pchl.i targ [flea.ADJZ market] FLEA MARKET – **Mod.ADJZ Head**

In the following sections, each of the nine binominal types is described and exemplified in more detail.[16]

3.2 The nine morphosyntactic strategies

3.2.1 No additional marking: jxt, cmp, der and cls

The ground floor (level 0) of the taxonomy contains four types, **jxt**, **cmp**, **der** and **cls**, each of which has exactly two components: the two thing-morphs that are the primary constituents of a binominal. There is no additional grammatical material.

3.2.1.1 jxt

The **jxt** ("juxtaposition") strategy involves a head and a modifier, both of which are thing-roots. There is no additional grammatical material and little or no fusion between the two constituents, which are treated as separate words.

The **jxt** type is found in 22% of the binominals in Pepper's database (Pepper 2020). It occurs in 76 of the 106 languages and is a significant word-formation type (accounting for at least 10% of binominals) in 53 of these. It accounts for the majority of binominals in 22 languages and is the only binominal word-formation strategy in Ceq Wong (CWG), Datooga (TCC), Imbabura Quechua (QVI), Seychelles Creole (CRS), Vietnamese and Walman (VAN).

Examples of the **jxt** strategy (all with the meaning RAILWAY) are given in (8):

16 All the examples, language names and glosses are as in Pepper (2020), where also the sources can be found

(8) a. Vietnamese *đường sắt* [road iron]
b. Saramaccan (SRM) *talán fútu* [train foot]
c. Western Farsi (PES) *rāh āhan* [way iron]
d. Kildin Sami (SJD) *rūvv't čuekas* [iron road]
e. Ho-Chunk (WIN) *mąąs nąągu* [metal road]
f. Cabécar (CJP) *kóbäkå̃ ña̱la̱* [train road]

3.2.1.2 cmp

The **cmp** ("compounding") strategy involves a head and a modifier, both of which are thing-roots. There is no additional grammatical material, but a high degree of fusion between the two constituents, such that the binominal constitutes a single word.

The **cmp** type is the most frequent type in Pepper's data set, accounting for almost 30% of all binominals. It occurs in 67 of the 106 languages and is a significant word-formation type (accounting for at least 10% of binominals) in 48 of these. Furthermore, it accounts for the majority of binominals in 24 languages and is the only binominal word-formation strategy in Caijia (caij1234) and Tuwari (TWW). The paradigm case of this type is the noun-noun compound, in which two nouns are simply concatenated. Examples of the **cmp** type (again, all with the meaning RAILWAY) are given in (9).

(9) a. Baa (KWB) *kràkísà* [road.train]
b. Bambara (BAM) *tɛrɛnsira* [train.road]
c. German *Eisenbahn* [iron.way]
d. Hawaiian (HAW) *alahao* [path.iron]
e. Mapudungun (ARN) *trenrüpü* [train.way]
f. Welsh (CYM) *rheilffordd* [rail.road]

Less prototypical examples of the **cmp** type are compounds that contain a linking element; these are discussed below in §4.1.4 in the context of gradient phenomena.

3.2.1.3 der

The **der** ("derivation") strategy involves a thing-root and a thing-affix. Less prototypically it can consist of two thing-affixes, as in neoclassical compounds.

The **der** type is found in just 432 of the 3,738 binominals in Pepper's database (12%), but it is attested in 60 of the 106 languages (57%). It is the preferred strategy in seven languages: Central Yupik (ESU), Croatian (HRV), Lithuanian (LIT), Oroqen (ORH), Polish, Puyuma (PYU) and Slovak, and the dominant strategy in two of these: Central Yupik and Puyuma.[17]

Some typical examples of the **der** type are listed in (10).

(10) a. Central Yupik *tallir.aq* [arm.AQ3] BRACELET
b. Puyuma *ka-ɭauk-an* [TMP-lunch-LOC] MIDDAY
c. Polish *wiatrak* [wind.NMLZ] WINDMILL
d. Czech *kůzle* [goat.DIM] KID
e. Hausa (HAU) *sàráunìyáa* [king.F] QUEEN
f. Gawwada (GWD) *sintitte* [nose.SG:F] NOSTRIL

Only affixes that contribute some tangible semantic content are considered in scope. The meaning contribution may be very general (THING, 10c) or it may be more specific (LOCATION 10b). Note that the gloss provides only a rough indication of the meaning contribution of the affix and is not claimed to be consistent. For one thing, the exact meaning of many derivational affixes is hard to pin down and may exhibit considerable variation; in addition, sources vary in terms of the degree of specificity used in glosses. As a case in point, Nagórko (2016: 2839) highlights the instrumental nature of the Polish suffix *-ak*, whereas it is glossed more generally as NMLZ in the binominals database (10c)

Diminutives are deemed to bear the meaning contribution SMALL THING. They can denote a small version of the entity denoted by the base (10d) or something small that is related in some way to the base entity (10a). Combinations of a thing-root and a gender-denoting affix are only considered to be binominals when the affix marks a clear semantic alternation. Thus, in (10e–f) QUEEN alternates with KING and NOSTRIL with NOSE. On the other hand, Gawwada *xarrap.atte* [spider_web.SG:F] SPIDER WEB is not regarded as a binominal since the suffix does not appear to derive a new meaning through gender alternation.

Neoclassical compounds constitute a non-prototypical variant of **der**. A word like *hydromancy*, is clearly a binominal but it consists of a prefix (*hydro-* 'water') and a suffix (*-mancy* 'divination'): in other words, of two thing-affixes.

17 Note, however, that the data set from Puyuma is very sparse and contains just three binominals, two of which are of type **der**. Most complex nouns in Puyuma, such as *pu-a-ɭima* [put-PJ-hand] GLOVE, have an actional component (often, as here, including a Projective Marker) and thus do not qualify as binominals.

3.2.1.4 cls

The **cls** ("classifier") strategy involves a thing-root and a noun classifier (thing-root or -affix). The denotatum of the binominal differs from that of the base, i.e., the classifier is used to derive a new meaning rather than for classification.

The **cls** type is the least frequent in Pepper's database, accounting for a mere 37 of the 3,738 binominals (1%). It is the preferred strategy in two of the 106 languages: Murui Huitoto (HUU, 11 out of 18 instances) and Trinitario (TRN, 12 out of 25). In addition it is found in Äiwoo (NFL), Bandial, Harakmbut and Swahili (SWA). It therefore requires much more detailed study. In order to facilitate further work, this section contains somewhat more detail than the others (see also Rose and Van linden, this volume, and Næss, this volume).

This type is motivated by the existence of forms such those in (11) and (12), which clearly qualify as binominals (examples from Urban 2012: 126–127).

(11) Arabela (ARL)
  a. *quitiaaca* [breast/teat.CL(liquid)] MILK
  b. *namijiaca* [eye.CL(liquid)] TEAR

(12) Bora
  a. *ííñuhéju* [earth.CL(hole)] CAVE
  b. *túúheju* [nose.CL(hole)] NOSTRIL

The classifier morphemes in these examples (*-aca* and *-héju*) have exactly the same function as the corresponding head constituents of, say, the Thai compounds, *náamtaa* [water.eye] TEAR and *ruucamùuk* [hole.nose] NOSTRIL. However, they cannot be used in isolation, so they are not thing-roots, and thus these binominals do not belong in the **cmp** type. These classifiers also constitute a closed class, which sets them off from the typical nominal constituents of **cmp** binominals. In both respects they are more like thing-affixes, so they could be classified under **der**. But classifiers differ markedly from affixes in having very precise semantics. This does not, in and of itself, constitute sufficient reason to separate them off from the **der** type, but the matter does not end there.

Aikhenvald (2000; 2017), citing criteria articulated earlier by Allan (1977), defines classifiers as "morphemes which occur in surface structures under specifiable conditions, denoting some salient semantic characteristics of the entity to which an associated noun refers". The examples from Arabela (11) and Bora (12) belong to one of seven subtypes of classifier in Aikhenvald's typology, which she

calls noun classifiers,[18] and are characterized by the fact that they "occur with a noun independently of any other constituent of a noun phrase or a clause". They can be affixes to nouns, as above, but they can also be "independent words with generic semantics" (13).

(13) Minangkabau (MIN) (Aikhenvald 2000)
  a. *batang limau* [CL(tree) lemon] LEMON-TREE
  b. *buah limau* [CL(fruit) lemon] LEMON-FRUIT

If the Arabela and Bora examples were to be classified as binominals of type **der**, then (13) must be classified as binominals of type **jxt**, and noun classifiers as a group would then be split across two binominal types. That is not necessarily a problem, but it suggests that a better solution – one that would make it possible to investigate the classifier phenomenon more closely – is to define a separate subtype **cls**.

The question is, how to define this type? Examples that parallel those from Minangkabau are also found in Atlantic-Congo languages. In (14) pairs of singular and plural noun class prefixes, *m-/mi-* and *Ø-/ma-*, distinguish trees from fruits, in just the same way as the Minangkabau classifiers *batang* and *buah*. If the Minangkabau words qualify as binominals, so too should the Swahili forms.

(14) Swahili (Russell 2003)
  a. *mlimau* / *milimau* [CL3.lemon / CL4.lemon] LEMON TREE/S
  b. *limau* / *malimau* [CL5:lemon / CL6.lemon] LEMON FRUIT/S

And so should the Bandial examples in (15), where the noun class prefixes serve, among much else, to distinguish between animals and their offspring (cf. the Czech diminutive suffix in 10d, above.

(15) Bandial (Watson 2015)
  a. *jijamen* [CL(ji).goat] KID
  b. *ejamen* [CL(e).goat] GOAT

Such noun class prefixes are not noun classifiers in Aikhenvald's typology. Instead they are classified under subtype (i) 'genders and noun classes' (see foot-

18 The seven subtypes identified by Aikhenvald are: (i) genders and noun classes, (ii) noun classifiers, (iii) numeral classifiers, (iv) classifiers in possessive constructions, (v) verbal classifiers, (vi) locative classifiers and (vii) deictic classifiers.

note 18 on page 17). One of the major differences between these two subtypes, according to Aikhenvald, is that in a noun class language every noun belongs to a noun class, whereas in noun classifier languages, every noun does not have to take a classifier.[19] Consequently, there would be a very substantial cost to admitting words like (15a, b) to the pantheon of binominals: Every noun in Swahili and Bandial would qualify as a binominal of type **cls** and, as a result, the data from noun class languages would swamp those from noun classifier languages and give a distorted overall impression of the **cls** type. That problem may not be insurmountable provided one remains aware of it, but unfortunately the issue is yet more complicated. Consider (16).

(16) Gawwada
  a. *pi?atte* [kid.SG:F] KID
  b. *xarrapatte* [spider_web.SG:F] SPIDER WEB

If the Bandial examples are regarded as binominals, why not also the Gawwada? After all, the only real difference between a two- or three-gender system (like the one in Gawwada and many Indo-European languages) and a noun class system of the Atlantic-Congo type is the size of the system: Aikhenvald groups them under the same subtype. And yet, the Gawwada examples cannot by any stretch of the imagination be regarded as the functional equivalents of noun-noun compounds. Moreover, recognizing them as binominals would lead to the kinds of construction we are interested in in this volume being completely lost from sight. Somewhere on this slippery slope a line has to be drawn.

That line could be drawn between Aikhenvald's two subtypes; it would amount to defining noun classifiers, but not noun class markers, as thing-morphs. (11)–(13) would then be categorized as binominals, while (14)–(16) would not. This would have the unfortunate consequence that (13) and (14), which really are parallel in every way, would be accorded different treatments. The line could also be drawn by contriving a distinction between noun class languages and gender languages based on the size of the system: say, more than three for noun class languages and two to three for gender languages. The line would then go between (15) and (16). Not only would this be somewhat arbitrary, it would also result in the aforementioned imbalance between noun class languages and noun classifier languages.

19 In addition, agreement is a necessary feature of noun class/gender systems but not of noun classifier systems. However, this does not impinge on the present discussion.

The solution adopted here is to draw the line instead between (15a) and (15b). The basis for making such a distinction is that in (15a) the denotatum of the whole (KID) is different from that of the base (GOAT), whereas in (15b) they are the same (GOAT and GOAT). In (15a) the noun classifier contributes a meaning component that *changes the denotatum*, i.e. its function is *derivational*. In (15b) this is not the case; nor is it in the two examples from Gawwada (16). Hence the qualification in the definition given above that the function of the classifier be derivational rather than classificatory.[20]

3.2.2 Marking on the head or modifier: prp, gen,adj and con

The first floor (level 1) of the binominal taxonomy also contains four types. What they have in common is that they contain one additional (grammatical) morpheme, over and above the two primary constituents. Three of the four (**prp**, **gen** and **adj**) share an apartment because the additional marker forms a constituent with the modifier; the fourth (**con**) lives alone, since the marker forms a constituent with the head.

3.2.2.1 prp

The **prp** ("prepositional") strategy involves a head and a modifier, both of which are thing-roots, and another independent lexeme that forms a constituent with the modifier.

The **prp** type accounts for 245 of the binominals in Pepper's database (6.5%), distributed across 27 languages, and it is the preferred strategy in eight of these: Barain, French, Italian (ITA), Maltese (MLT), Romanian (RON), Swahili, Tagalog (TGL) and Tarifit (RIF).

In the typical case, exemplified in (17a–c), the additional lexeme is a preposition (hence the choice of mnemonic for this type), but it may also be a postposition (17d) or a particle named according to a language-specific descriptive category, such as a connector (17e) or linker (17f).

(17) a. French *chemin de fer* [road of iron] RAILWAY
b. French *moulin à vent* [mill to wind] WINDMILL
c. Welsh *papur lle chwech* [paper for toilet] TOILET PAPER

20 See Pepper (2020: 148–154) and Rose and Van linden (this volume) for further discussion.

d. Hindi (HIN) *dāṃt kā braś* [tooth GEN brush] TOOTHBRUSH
e. Lingala *nzela ya masini** [way CON train] RAILWAY (* no relation)
f. Tagalog *butas ng ilong* [hole LK nose] NOSTRIL

While prepositions are fairly common, postpositions are rare and can be problematic, in that they can often be analysed as case affixes (i.e. **gen**) rather than adpositions (**prp**). The Hindi example (17d) is a case in point. It is glossed using the abbreviation GEN in the database (as decided by the contributor), but *ka* is also commonly regarded as a postposition (§4.1.5).

The most commonly used adpositions are those whose function also includes to indicate possession, such as the French *de* (17a) and the Hindi *ka* (17d), but some languages permit other prepositions to be used as well, such as a locative (17b) or purposive (17c). In other languages, the particle has a more general, associative meaning that is used for a wide variety of relations and not solely for possession. Examples include the Lingala Connective *-a* (17e) and the Tagalog Linker *ng* (17f).

In French, more than one preposition is available for use in binominal word-formation (17a-b). This suggests that the present typology may be too coarse-grained for certain kinds of investigation, for example, into the semantics of French prepositional compounds – or that it should be used in conjunction with a classification of semantic relations like Hatcher-Bourque (see Pepper, b, this volume).

3.2.2.2 gen

The **gen** ("genitival") strategy involves a head and a modifier, both of them thing-roots, with an additional, non-transpositional affix or segmental marker attached to the modifier.

There are 484 instances of the **gen** type in Pepper's database (13%), making it the third most frequent strategy in absolute terms (after **cmp** and **jxt**). It is also ranked third in terms of the number of languages in which it is found (55 out of 106, i.e. 52%). It is the preferred strategy in 15 of those languages: Amharic (AMH), Archi (AQC), Assamese (ASM), Bezhta, Estonian (EST), Gawwada, Greek (ELL), Irish (GLE), Kambaata (KTB), Kanuri (KAU), Latvian (LAV), Nepali (NEP), Sidamo (SID), Wawa (WWW) and Zinacantán Tzotzil (TZO).

Typically, the additional affix indicates the genitive case (18a-c) or possessive function (18d), but other cases occur as well, including the dative (18e).

(18) a. Bezhta *kilos hino* [iron.GEN road] RAILWAY
b. Irish *muileann gaoithe* [mill wind:GEN] WINDMILL
c. Kanuri *súwúlí kánzàbè* [opening nose.GEN] NOSTRIL
d. Takia (TBC) *graŋen tatu* [side.3SG bone] RIB
e. Gurinji *yawartawu marru* [horse.DAT house] STABLE OR STALL
f. Tarifit *tisi ufus* [bottom STC.hand] PALM OF HAND

The Tarifit example (18f) illustrates the kind of confusion that arises if one assumes that descriptive categories are the same across languages. Here the *modifier*, 'hand', normally *fus*, is in what some Berber linguists call the "construct state" (hence the gloss, *status constructus*). The very same term is used in Semitic linguistics to describe a special form of the *head* in adnominal constructions. Consequently, Berber words glossed with STC belong to the type **gen** (dependent-marked) whereas Semitic words glossed with STC belong to the type **con** (head-marked), cf. (23e) below.

3.2.2.3 adj

The **adj** ("adjectival") strategy involves a head and a modifier, both of them thing-roots, with an additional, transpositional morpheme attached to the modifier.

The binominals database contains 196 instances of the type **adj** (5%) and it occurs in 28 languages (26%). The great majority of these are European languages, either Indo-European or Uralic (19). The six Slavic languages (Croatian, Czech, Lower Sorbian, Polish, Russian and Slovak) account for 130 of them alone. Whether this is because adjective as a productive lexical category is more frequent in Europe than elsewhere is a question for further research.

(19) a. Italian *via lattea* [way milk.ADJZ] MILKY WAY
b. Lithuanian *geležinkelis* [iron.ADJZ.way] RAILWAY
c. Polish *kolej żelazna* [course iron.ADJZ] RAILWAY (arch.)
d. Polish *złoty pierścionek* [gold.ADJZ ring] GOLD RING
e. Russian *železnaja doroga* [iron.ADJZ road] RAILWAY
f. Kildin Sami *mājjtjes' lījjhm* [milk.ATTR cow] DAIRY COW
g. Hungarian (HUN) *északi fény* [north.ADJZ light] ARCTIC LIGHTS
h. Hungarian *képeslap* [picture.PROP.card] POSTCARD

The most common descriptive category for the additional morpheme is adjectivizer, but the terms attributivizer and proprietive are also found. Polish and Hun-

garian are notable for having two distinct constructions of this type. In Polish the same construction can be either head initial (19c) or head-final (19d). In Hungarian there are two different adjectival suffixes (19g-h): *-i* (labelled ADJZ) and *-s* (labelled PROP), both of which can be attached to a wide variety of nouns (Kiefer 2009).

Other examples are found scattered across the globe in Africa (20a), the Caucasus (20b), Asia (20c-e), New Guinea (20f) and Central America (20g).

(20) a. Kanuri *kámú nyìyáà* [woman marriage.ADJZ] MARRIED WOMAN
b. Bezhta *kaǰ'ako tormoz* [hand.OBL.SUP.ATTR brake] HAND BRAKE
c. Western Farsi *asiyāb bādi* [mill wind.ADJZ] WINDMILL
d. Ket *soltu təqol* [gold.ADJZ finger.covering] GOLD RING
e. Yakut (SAH) *tualetnay kuma:yi* [toilet.ADJZ paper] TOILET PAPER
f. Kalamang (KGV) *sontum warten* [person sorcery.ADJZ] SORCERER OR WITCH
g. Kekchí (KEK) *k'imal kab'l* [straw.ADJZ house] THATCH

3.2.2.4 con

The **con** ("construct") strategy involves a head and a modifier, both of them thing-roots, with an additional, non-transpositional affix or segmental marker on the head.

The type **con** accounts for 351 of the binominals in Pepper's database (9%) and is found in 24 of the 106 languages (23%). It is the preferred strategy in 10 of these: Anindilyakwa (AOI), Hausa, Hebrew (HEB), Iraqw (IRK), Kekchí, Kupsabiny (KPZ), Turkish, Western Farsi, Wolof (WOL) and Yakut; and it accounts for over 75% of all binominals in Hausa, Hebrew, Kupsabiny and Wolof.

The term 'construct' is traditionally used in Semitic linguistics but has been extended by Creissels (2017; this volume) to cover any obligatory marking on a noun that fulfils the role of head in nominal modifier constructions, provided it does not cross-reference features of the modifier that condition its use. Because of the latter proviso, example (22b) is a clear case of a binominal of type **con** that is not an instance of construct marking according to his definition, "since *x-* is an uncontroversial 3rd person singular prefix" (p.c.). Apart from this proviso, Creissel's term 'construct' covers every instance in the binominal database of the type **con**.

Binominals of this type are glossed in a variety of ways (21–23). Labels used in traditions other than Semitic include linker, possessive, genitive and pertensive.[21]

21 In addition, Haspelmath (2009) proposes the term "anti-genitive".

The latter term, proposed by Dixon (2010b), is restricted to possessive constructions (unlike Creissels' construct), but does permit cross-referencing.

The type **con** covers what Croft (2003; 2022) terms 'linkers' and 'special forms' (to the extent that these occur on the head), as well as his 'indexical' strategies. Croft subdivides the latter according to whether or not they encode the category of person, into 'person indexation' and 'nonperson indexation'. Koptjevskaja-Tamm (2003: 645) makes what appears to be a similar distinction, albeit using different terminology, between two subtypes of head-marking:

1. **relators**, whereby the form of the head signals the presence of the dependent in the same NP, without, however, specifying its features;
2. **indexers**, whereby the form of the head varies according to the properties of the dependent.

Koptjevskaja-Tamm's relators correspond to Croft's linkers and special forms, since they do not exhibit contrast, whereas her indexers may involve either person indexation or nonperson indexation. Examples of the latter are found in Barain (21a-b) and Hausa (21c-d), where the markers are *-ji*/*-(g)eti* and *-r*/*-n*, respectively, depending on the gender of the dependent.

(21) a. Barain *sinja guma-geti* [nose$_{F}$ hole-POSS:3SG:F] NOSTRIL
b. Barain *nokuno non-ji* [goat$_{M}$ child-POSS:3SG:M] KID
c. Hausa *kàfá-r háncìi* [orifice$_{F}$-LK nose] NOSTRIL
d. Hausa *dóokì-n kárfèe* [horse$_{M}$-LK metal] BICYCLE

On the other hand, the glossing of (22a-d), all of which make reference to the third person, suggests that these are indexers in Koptjevskaja-Tamm's scheme and examples of person indexation in Croft's.

(22) a. Kalamang *kanggir pul-un* [eye skin-3POSS] EYELID
b. Kekchí *x-na'aj xam* [3ERG-place fire] FIREPLACE
c. Takia *su mala-n* [breast eye-3SG] NIPPLE OR TEAT
d. Yakut *χaraχ u:-ta* [eye water-3SG] TEAR

Lastly, the invariant possessive affixes in Kupsabiny (23a) and Malagasy (PLT) (23b) are relators for Koptjevskaja-Tamm but linkers for Croft. So too are the Galibi Carib (CAR) possessive suffix *-li* and its allomorph *-yi* (23c-d) since their distributions are phonologically determined and not conditioned by features of either the head or the modifier. The form of the Hebrew construct case (23e) is determined by the gender of the head, so it is a Koptjevskaja-Tamm relator, whereas it is a

special form for Croft. Despite this variation, I regard all of the examples in this section as binominals of type **con**.

(23) a. Kupsabiny *kariit-aap maata* [car-POSS fire] TRAIN
b. Malagasy *lala-m-by* [road-PER-iron] RAILWAY
c. Galibi Carib *manatɨ potɨ-lɨ* [breast tip-POSS] NIPPLE OR TEAT
d. Galibi Carib *upupo kuwai-yɨ* [head calabash-POSS] SKULL
e. Hebrew *mesila-t barzel* [track-STC iron] RAILWAY

3.2.3 Marking on the head and modifier: dbl

In the attic (level 2) of the binominal house there is just one type: **dbl**. Like most attics, the contents are somewhat untidy, as will be explained shortly.

3.2.3.1 dbl

The **dbl** ("double") strategy involves a head and a modifier, both of them thing-roots, with additional morphemes attached to both.

There are just 37 instances of the **dbl** type in Pepper's database, a mere 1% of the total. They are distributed across 15 languages: Akkadian (AKK), Barain, Central Yupik, Galibi Carib, Hebrew, Kekchí, Maltese, Oroqen, Puyuma, Romanian, Seri, Somali (SOM), Takia, Trinitario and Western Farsi, in four of which they are the preferred binominal strategy: Akkadian, Seri (SEI), Somali and Takia.

(24) a. Galibi Carib *emo-lɨ sakɨla-lɨ* [nose-POSS aperture-POSS] NOSTRIL
b. Takia *patu-n kdabog-an* [egg-3SG yellow-3SG] YOLK
c. Oroqen *dalay-ŋi ŋəːkə-n* [sea-GEN bank-3SG:POSS] SHORE
d. Somali *bam-ka biyo-ha* [pump-DEF water-DEF] WATER PUMP
e. Maltese *l-isfar tal-bajda* [DEF-yellow of:DEF-egg] YOLK
f. Akkadian *piliš app-im* [hole:STC nose-GEN] NOSTRIL
g. Kambaata *qiissann-a wodar-u* [spider-F:GEN line-M:GEN] SPIDER WEB

As the examples in (24) demonstrate, there is considerable variation in terms of the kinds of markers (case, definiteness, possession, construct), and the ways in which they are combined. Sometimes it is the same affix that attaches to both major constituents (24a-b). In some languages the markers appear to cross-reference each other (24b), in others the affix on the head cross-references the mod-

ifier (24c). Somali exhibits two definiteness markers (24d) and Maltese a combination of definite marker and definite preposition (24e). Finally, Akkadian (24f) exhibits an older form of the Semitic construct state with the modifier in the oblique case, while in Kambaata (24g) both elements have genitive markers. The variety encountered here suggests that a more fine-grained classification would be possible. However, the binominals database contains too little data for this to be feasible. With more data these could be analysed in terms of combinations (one from each apartment) of the types found on level 1, and perhaps also Croft's distinction between relators, indexes and linkers.

3.2.4 Summary of binominal strategies

For ease of reference we conclude this section with a summary table of the nine binominal strategies, with their mnemonics, definitions and examples (of either RAILWAY or NOSTRIL).

3.3 Unattested strategies

As noted above and shown in Figure 2 on page 37), level 1 of the classification is divided into two "apartments", with dependent-marking strategies to the left and head-marking strategies to the right. The three dependent-marking strategies (**prp**, **gen** and **adj**) are situated from right to left, in that order, such as to reflect Bybee's (1985) scale based on degree of fusion. The single head-marking strategy (**con**) is situated in the middle of the right-hand section in order to highlight its symmetrical relation to **gen**, since **gen** is a non-transpositional affixing strategy associated with the dependent, and **con** is a non-transpositional affixing strategy associated with the head. Once the nine types are laid out in this manner, two gaps are revealed, labelled (**prn**) and (**nml**). These are the head-marking correlates of **prp** and **adj**, respectively. In this section we discuss possible explanations for these lacunae.

3.3.1 (prn)

The first missing type is the head-marking correlate of **prp**, which we have labelled (**prn**) for reasons that will become apparent. For such a type to exist, it must consist of a head, a modifier and another independent lexeme that forms a co-constituent with the head, e.g. **Mod {X Head}**.

Table 4: Summary of binominal strategies.

Mnemonic	Markers	Definition	Example
jxt	**0**	The **jxt** ("juxtaposition") strategy involves a head and a modifier, both of which are thing-roots. There is no additional grammatical material and little or no fusion between the two constituents, which are treated as separate words.	Saramaccan (SRM) *talán fútu* [train foot] RAILWAY
cmp		The **cmp** ("compounding") strategy involves a head and a modifier, both of which are thing-roots. There is no additional grammatical material, but a high degree of fusion between the two constituents, such that the binominal constitutes a single word.	German (DEU) *Eisenbahn* [iron.way] RAILWAY
der		The **der** ("derivation") strategy involves a thing-root and a thing-affix. Less prototypically it can consist of two thing-affixes, as in neoclassical compounds.	Gawwada (GWD) *sintitte* [nose.SG:F] NOSTRIL
cls		The **cls** ("classifier") strategy involves a thing-root and a noun classifier (thing-root or-affix). The denotatum of the binominal differs from that of the base; i.e., the classifier is used to derive a new meaning rather than for classification.	Bora (BOA) *túúheju* [nose.CL(hole)] NOSTRIL
prp	**1 (modifier)**	The **prp** ("prepositional") strategy involves a head and a modifier, both of which are thing-roots, and another independent lexeme that forms a constituent with the modifier.	French (FRA) *chemin de fer* [road of iron] RAILWAY
gen		The **gen** ("genitival") strategy involves a head and a modifier, both of them thing-roots, with an additional, non-transpositional affix or segmental marker attached to the modifier.	Amharic (AMH) *yebaburi ḫādīdi* [GEN.train way] RAILWAY
adj		The **adj** ("adjectival") strategy involves a head and a modifier, both of them thing-roots, with an additional, transpositional morpheme attached to the modifier.	Russian (RUS) *železnaja doroga* [iron.ADJZ road] RAILWAY
con	**1 (head)**	The **con** ("construct") strategy involves a head and a modifier, both of them thing-roots, with an additional, non-transpositional affix or segmental marker on the head.	Malagasy (PLT) *lalamby* [road.PER.iron] RAILWAY
dbl	**2 (head and modifier)**	The **dbl** ("double") strategy involves a head and a modifier, both of them thing-roots, with additional morphemes attached to both.	Akkadian (AKK) *piliš appim* [hole:STC nose.GEN] NOSTRIL

So what kind of item might be a candidate for the role of X? One way to approach this question is to look for a relation **prn** ↔ **con** that is isomorphic with the relation **prp** ↔ **gen**. Now, it is well-established that adpositions (**prp**) are a common source of case markers (**gen**): "Diachronically, case affixes arise from adpositions that become affixed to the noun" (Croft 1990: 34). The missing type **prn** could thus be whatever is the source of **con**.

According to Croft (2003: 35–36), "bound indexation markers", such as those in binominals of type **con**, develop out of (i) pronouns (in the case of person indexation) and (ii) articles (in the case of nonperson indexation). An example of the former is the Hausa (construct state) suffix *-n* (plural or masculine singular) or *-ɾ̃* (feminine singular), which attaches to the head in possessive constructions (25a, c). This suffix also occurs in the **Head.LK Mod** construction responsible for 40 of the 43 Hausa binominals in the database used in the present study (cf. examples 21c-d on page 22). According to Creissels (2009), this suffix results from the cliticization of a pronoun *na*/*ta* that is co-referent with the head noun in the synonymous construction illustrated by (25b, d).[22]

(25) Hausa (cf. *kàree* 'dog', *saanìyaa* 'cow')
  a. *kàre.n Daudà* [dog_{M}.CSTR:SG:M Dauda] 'Dauda's dog'
  b. *kàree na Daudà* [dog_{M} that one (SG:M) of Dauda] 'Dauda's dog'
  c. *saanìya.ɾ̃ Daudà* [cow_{F}.CSTR:SG:F Dauda] 'Dauda's cow'
  d. *saanìyaa ta Daudà* [cow_{F} that one (SG:F) of Dauda] 'Dauda's cow'

In other words, the source of **Head.LK Mod** is **Head, PRON Mod**. The latter construction would be considered an instance of the missing type **prn** if the pronoun formed a constituent with the head (i.e., {**Head PRON**} **Mod**), but that is not the case. Instead, the pronoun forms a constituent with the modifier (**Head** {**PRON Mod**}), which means that (25b) and (25d), if they were binominals (which they are not, because they do not have a naming function), would be instances of **prp**, not **prn**. This is an example of reanalysis, in which an element preposed to the modifier in a head-initial construction is reinterpreted as a postposed marker on the head (26).

(26) *kàree {na Daudà}* Head, PRON Mod ➔ {*kàre.n} Daudà* Head.LK Mod

22 Newman (2000: 300) calls *na*/*ta* a (free) (genitive) linker. It can combine with personal pronouns, but is not itself a pronoun, according to him.

Clearly, constituency must be taken into consideration when looking for examples of the missing type **prn**. There are six logical possibilities (27). The component X might be a pronoun or an article, but it must form a constituent with the head. This means that constructions (27c) and (27d) are highly unlikely: they could only occur in a non-configurational language.

(27) a. **{X Head} Mod** b. **{Head X} Mod** c. ~~Head {Mod} X~~
d. ~~X {Mod} Head~~ e. **Mod {X Head}** f. **Mod {Head X}**

An example of (e), in which a pronoun copy of the dependent (*ha* 'he') is preposed to the head, is provided by Nichols (1992: 79) from Atakapa (AQP, extinct isolate), together with a schematic English rendition (28a,b).

(28) a. *yukhiti icak kau ha tal*
Indian man dead he skin
'the skin of a dead Atakapa'
b. [[the man] [he skin]] 'the man's skin'

(28a) is, of course, a possessive construction, not a binominal, so it does not count as an instance of the missing binominal type **prn**. However, since binominals often recruit their morphosyntactic strategies from possessive constructions (see below), it is perfectly possible that the type does exist somewhere. Were it to be found, it would correspond to Koptjevskaja-Tamm's "linking pronoun", the one type in her PNP classification that was not found in Pepper's binominal data. While the linking pronoun type of PNP is rare in European languages, its status across the world's language is unknown and it seems eminently possible that the binominal type **prn** could exist somewhere. Finding it, however, must remain a topic for further research.

3.3.2 (nml)

The second missing type is the head-marking correlate of **adj**, labelled **nml**. If such a type exists, it must consist of a head, a modifier and a transpositional (i.e., word class changing) morpheme attached to the head, just as **adj** consists of a head, a modifier and a transpositional morpheme attached to the modifier, cf. the Russian example *železnaja doroga* [iron.ADJZ road] RAILWAY. There are two logical possibilities:

– Either the additional morpheme is a nominalizer – that is to say, **Mod Head. NMLZ**, in which case the head element would not be a thing-root;

– or it derives some other word class – as for example in **Mod Head.ADJZ**, in which case the resulting construction would not denote an object, but rather a property.

In neither case would the form in question be regarded as a binominal. In other words, **nml** as a type of binominal appears to be a logical impossibility, at least as long as one thinks in terms of major word classes; it is not found in the data for a good reason.

Having said that, van Egmond (this volume) proposes the INALP construction in Anindilyakwa as a possible candidate for the **nml** type. Her argument is that the Anindilyakwa Inalienable Possession (INALP) construction is used to denote

> parts of inanimate objects, plants, animals, and to components of body parts. The 'part' noun [i.e. the head] is marked for INALP and maintains its intrinsic noun class prefix, as this is frozen to the stem. The derived nominal behaves like an adjective in that it is now flexible and can take any pronominal/gender/noun class prefix to agree with the independent noun [i.e. the modifier] that represents the 'whole'. (page 164)

Van Egmond represents the structure as in (29a) and provides examples like (29b):

(29) a. [NCx-(G-)INALP-NCy.Head (NCx.Mod)]
b. *ma-m-ayarrka mukayuwa*
VEG-INALP-NEUT.hand VEG.dillybag
'handle of dillybag'

In (29b), *ayarrka* 'hand' is used metonymically to denote a handle and is thus the head of a construction in which the modifier is *mukayuwa* 'dillybag'. Both of these are thing-morphs, and there is no actional element, so this is clearly a binominal. Now, *ayarrka* belongs to the neuter noun class and *mukayuwa* to the vegetable noun class. They cannot simply be juxtaposed because there is obligatory agreement throughout the clause in Anindilyakwa. So what happens is that an INALP prefix *m(a)-* is added to the head constituent, thereby permitting the VEG noun class prefix required by the modifier (*ma-*) to be attached to the head.

In van Egmond's analysis, this is tantamount to changing *ayarrka* into an Anindilyakwa Adjective (a property word); the process is transpositional (word-class changing) and since the marker is associated with the head, we therefore have an instance of the missing type **nml**. However, it is not the case that *ayarrka* ('hand') qualifies *mukayuwa* ('dillybag'); it is the reverse. The situation is thus no different in principle from the Hausa examples (21c-d) in which the form of the morpheme attached to the head is governed by the gender of the modifier. Like

the Hausa binominals, the Anindilyakwa Inalienable Possession construction must therefore be classified as **con** in the binominal typology proposed here.

In conclusion, binominals consisting of two thing-roots and a transpositional morpheme attached to the head (i.e. **nml**) have not yet been discovered, and given the current definition, they would appear to be a logical impossibility.

4 Binominals and grammaticalization

4.1 Grammaticalization pathways

The two-dimensional representation of the typology of binominal lexemes was developed in order to account for gradient phenomena. This section discusses instances of constructions that fall in between the nine major types. It is based primarily on the data collected by Pepper (2020) but includes some examples from other sources. Each subsection refers to one of the numbered items in Figure 4.

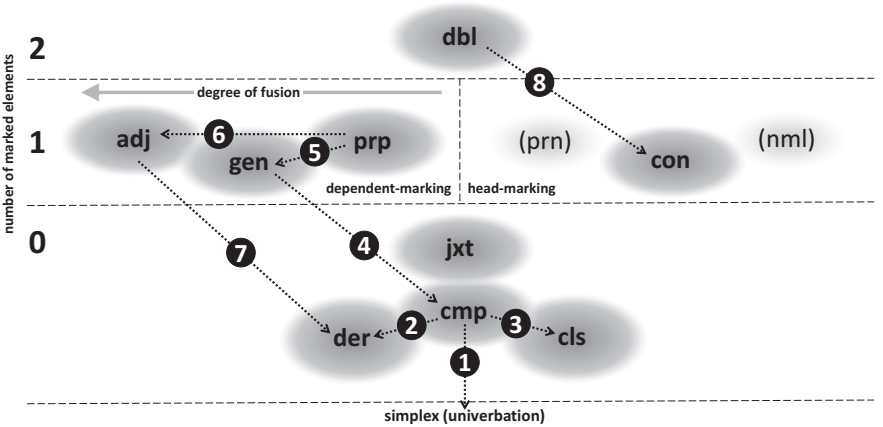

Figure 4: Formal classification showing gradient phenomena.

4.1.1 Univerbation (cmp ➔ simplex) ❶

Univerbation is the term given to the historical process by which an (analysable) item consisting of two or more morphemes develops into an (unanalysable) item consisting of a single morpheme. Examples of such simplex forms belong in the

basement of the binominal house, i.e. level -1, but since they would not be binominals (by definition), such a level is not required for the classification.

However, since univerbation is a gradual process, it is only to be expected that there will be partially analysable items that are intermediate between the types on level 0 (**der**, **jxt**, **cmp**, **cls**) and that lower level. Two examples are given in (30). Eng. *nostril* (30a) was originally a compound but is no longer identifiable as such, despite the first constituent being recognisable as 'nose'. In Eng. *lord* (30b) and Ger. *Messer* (30c), on the other hand, the process of univerbation has reached its end-point: What started out as prototypical binominals of type **cmp** are today completely opaque to lay speakers.

(30) a. Eng. *nostril* < *nose* + *thirl* ('hole')
b. Eng. *lord* < *hlāf* ('bread') + *weard* ('guardian')
c. Ger. *Messer* 'knife' < Proto-Ger. **matiz* 'food' + **sahsą* 'knife, dagger'

4.1.2 Affixoids (cmp ➔ der) ❷

The difference between **cmp** and **der** is that the former consists of two thing-roots whereas the latter consists of a thing-root and a thing-affix. However, the distinction between a root and an affix is not clear-cut; the two exist as end-points on a continuum that can be defined in terms of autonomy versus dependence (Tuggy 1992). Between these two end-points one finds phenomena called affixoids that are neither fully autonomous nor fully dependent. Booij (2010) gives a number of examples from Dutch in which a noun acquires a specialized meaning when used as the head of a compound (31).

(31) *baron* 'baron' > rich dealer: *afval-baron* [trash-baron] 'rich dealer in trash'
boer 'farmer' > seller: *sigaren-boer* [cigar-farmer] 'cigar seller'
man 'man' > seller: *bladen-man* [magazine-man] 'magazine seller'
marathon 'marathon' > long-session: *jazz-marathon* 'jazz marathon'

In English *postman* and many other compounds whose second constituent is *man*, phonological reduction of the second element *-man* from /mæn/ to /mən/ indicates a status intermediate between root and affix, even though it may not yet have "broken away from MAN, becoming a lexical formative on its own" (Matthews 1991: 94).

4.1.3 Bound nouns (cmp ➔ cls) ❸

As pointed out in §3.2.1.4, the type **cls** ("classifier") is the least well-defined and the most poorly represented in the database. If terminology is anything to go by, it consists of a number of somewhat disparate phenomena, as witnessed by the many transitional cases in Äiwoo (Næss, this volume). One of the strategies found in this language uses a "bound noun", a term suggestive of something intermediate between a noun and a classifier, which would in turn give rise to binominals mid-way between **cmp** and **cls**. The distinction between classifiers and bound nouns is discussed in detail by Rose and Van linden (this volume) in their description of the Western Amazonian languages Trinitario and Harakmbut, and they note "the analytical problem" of distinguishing between the types **cmp** and **cls**.

4.1.4 Linking elements (gen ➔ cmp) ❹

In many languages noun-noun compounds involve linking elements. Almost all the examples in Pepper's database are from Indo-European languages (32a-d), the only exceptions being from Korean (KOR) (32e), where it is found in what Yeon & Brown (2011) describe as "compounds in which the two elements are linked together by the addition of the so-called 'genitive s'" (p. 31). The latter, which causes tensing (or reinforcement) on the following plain consonant, is best regarded as a linking element in the modern language. Bauer (2001) cites an example from Cambodian (KHM) (32f) and mentions Yoruba (YOR) as having a "purely phonological" linking element that involves prolongation of the final vowel, while W. Bauer (1993) mentions a type of compounding involving a linking element *-aa-* "which is being increasingly used at present" in Maori (MRI) (32g).

(32) a. German *Nasenloch* [nose.LE.hole] NOSTRIL
b. Greek *siδiroδromos* [iron.LE.road] RAILWAY
c. Lithuanian *voratinklis* [spider.LE.web] SPIDER WEB
d. Russian *golenostop* [shank.LE.foot] ANKLE
e. Korean *khoskwumeng* [nose.GEN.hole] NOSTRIL
f. Cambodian *yianəthaan* [vehicle.LK.place] GARAGE
g. Maori *waiata-aa-ringa* [song-LK-hand] ACTION SONG

Many elements of this kind have their origin in case and/or number suffixes that have become semantically bleached and now often conflict with the grammar. For example, in the German *Regierungschef* [government.LE.head] 'head of govern-

ment' the linking element *-s-*, a reflex of the masculine genitive, is here attached to a feminine noun. The Greek linking element *-o-* (32b) originates in an ancient thematic vowel but today functions solely as a compounding marker (Ralli 2013). Binominals such as these can be said to occupy the space between **gen** and **cmp** but are arguably closer to the latter. Other linking elements, like those in Yoruba and Cambodian, may only ever have had a phonological role.

Binominals with linking elements thus present a challenge when coding the data: classifying them consistently as either **cmp** or **gen** could obscure important distinctions in Germanic and Greek respectively. The solution adopted by Pepper (2020: 166) is to classify them in such a way as to bring out any contrasts that might be relevant in each individual language. Thus Germanic binominals with linking elements are coded as **gen** (to contrast with the **cmp** strategy otherwise found in those languages) and as **cmp** in Greek (to contrast with the "true" **gen** strategy).

4.1.5 Adpositions or case affixes? (prp ➔ gen) ❺

As noted above (§3.3.1), case affixes often arise from adpositions that become attached to the noun. As a result, the status of some binominals as either **prp** or **gen** can be hard to determine. A classic example is the Japanese *no* construction which some linguists analyse as a genitive suffix (33a) and others as a postposition (33b). The orthography offers no clue since the particle *no* is written in Hiragana (の) while the other words are written in Kanji (蜘蛛の巣). In order to facilitate comparison with Korean, in which the equivalent possessive particle 의 (*-uy*) is always written as a suffix, Pepper took the decision to classify the Japanese forms as **gen** rather than **prp**.

(33) Japanese
  a. *kumonosu* [spider.GEN.web] SPIDER WEB
  b. *kumo no su* [spider POSTP web] SPIDER WEB

The orthography used in Maltese, on the other hand (34), suggests that the combination of the preposition *ta'* and the definite article *il-*, which occurs commonly in binominals, is neither a separate word nor a prefix, but rather a clitic. This, again, lies somewhere between **prp** and **gen**.

(34) Maltese *mitħna tar-riħ* [mill OF:DEF-wind] WINDMILL

Sometimes grammatical descriptions analyse equivalent constructions in closely related languages in rather different ways. This applies to possessive constructions in Hindi and Nepali. Whereas in Hindi (35a) the possessive morpheme is written, transliterated and referred to as a postposition, in Nepali (35b) it is treated as a suffix. The decision as to which category to assign must be taken in such a way as to minimize any adverse analytical consequences.

(35) a. Hindi मकड़ी का जाला *makṛī **kā** jālā* [spider POSTP web] SPIDER WEB
b. Nepali माकुराको जालो *mākurā.**ko** jālo* [spider.GEN web] SPIDER WEB

4.1.6 Inflection or derivation? (gen ~ adj) ❺

The definitions of **gen** and **adj** in (§3.2.2.2 and §3.2.2.3) do not make reference to the notions of inflection and derivation, but rather to the distinction between transpositional (word-class changing) and non-transpositional affixation. The reason for this is that the traditional distinction between inflection and derivation, whereby derivational affixes change the word-class of their base, while inflectional affixes do not, has been shown to be too simplistic. Haspelmath (1996) uses the example of Slavic possessive adjectives to show that the difference between inflection and derivation is one of degree, with Upper Sorbian being at the inflectional end of the scale and Russian more towards the derivational end. Defining **gen** and **adj** in terms of inflection and derivation would thus result in intermediate forms. Defining them in terms of transposition, on the other hand, results in a more clear-cut distinction.

4.1.7 Head replacement (adj ➔ der) ❼

The type **adj** belongs to level 1 in the classification whereas **der** belongs on level 0: the former has three components whereas the latter has just two. An intermediate between these two is represented by the Slovak word *železnica* (36a). The structure of this word parallels that of the Russian *železnaja doroga* (36b) precisely, except for the use of the nominalizing suffix *-ica* instead of a lexical head *doroga*.

(36) a. Slovak *žel*ez.n.ica* [iron.ADJZ.NMLZ] RAILWAY
b. Russian *žel*ez.naja doroga* [iron.ADJZ road] RAILWAY

Thus in one sense the word belongs on level 1 under **adj**. On the other hand, as a derived word it has more in common with other derivations and, indeed, Slovak linguists recognize an alternative synchronic analysis, *žel*ez-*nica* [iron-NMLZ], an undoubted instance of the **der** type:

There are two possible starting points for the analysis of the word *železnica*:

1. It is derived from *železo (iron)* and can be paraphrased as follows: "the object which is related to iron" (which moves on iron)
2. It is derived from *železný* (iron_{ADJ}) as univerbization from *železná dráha* ('railway').

(Martina Ivanova, p.c. via Lívia Körtvélyessy)

This form can thus be seen as intermediate between **adj** and **der** and represents a type that occurs rather often in certain Slavic languages, in which the head element of an adjectival binominal is replaced by a more general nominalizing suffix.

4.1.8 Morpheme loss (dbl ➔ con) ❽

The final example of intermediate (gradient) phenomena is that of morpheme loss. Citing data from Hungarian, Kirmandji (KMR), Arbore (ARV) and Maltese, Koptjevskaja-Tamm (2003) shows that "the step between double-marking and head-marking [in PNPs] is not necessarily big":

> Head-marked PNPs in Maltese, similarly to head-marked PNPs in Kirmandji, have developed from earlier double-marked PNPs, partly due to the breakdown of the case system of modern Arabic dialects compared to Classical Arabic, in which the possessor regularly appeared in the genitive case. (p. 647)

The same appears to be the case with binominals, and not just between **dbl** and **con** (the example shown in Figure 4), but also between **dbl** and **gen**, between **prp** and **gen**, and between **gen** and **con** on the one hand and **cmp** on the other. Or more generally, between any strategy involving *n* additional morphemes and one involving *n-1* morphemes.

One particularly striking example is Welsh, in which the dominant type at an earlier stage of the language was **gen** (as it still is in Irish). Following the loss of case marking the dominant types are today **jxt** and **cmp**. Elsewhere in the database there are indications that this process is at work in Galibi Carib, Tarifit and Swahili (37)-(39). In the case of Galibi Carib, examples (37a-c) are **dbl**, **gen** and **con**, respectively. The double-marked pattern (a) may represent an earlier construction from which the others have developed.

(37) Galibi Carib
  a. *emo.li sakila.li* [nose.POSS aperture.POSS] NOSTRIL
  b. *pana.li weti* [ear.POSS dirtiness] EARWAX
  c. *manati poti.li* [breast tip.POSS] NIPPLE OR TEAT

Example (38) is one of three words in the Tarifit sub-vocabulary of Pepper's database in which the preposition *n* is given as optional. With the preposition the construction is considered an instance of **prp**; without it, it is an instance of **gen**.

(38) Tarifit *tisi (n) ufus* [bottom (of) hand:STC] PALM OF HAND

There is also an example in Swahili (39) of a construction in which the associative marker is given as optional. Since it is the only occurrence, it is classified as **prp** by Pepper along with other words that exhibit this marker, but it may also indicate gradience.

(39) Swahili *gari (la) moshi* [car (CON) smoke] TRAIN

4.2 Binominals and adnominal possessives

4.2.1 The modification-reference continuum

Intuitively, as was suggested in §3.1, binominals are closely related to adnominal possessive constructions, and this is demonstrated by the degree of overlap between the typological classification of binominals presented in this chapter and Koptjevskaja-Tamm's classification of possessive noun phrases in Europe. The two are almost identical: seven of Koptjevskaja-Tamm's eight types have been documented for binominals, and the eighth (**prn**), it has been suggested (§3.3.1), is probably out there somewhere, waiting to be discovered. Just two types needed to be added to Koptjevskaja-Tamm's classification (**der** and **cls**) and the former would have to be added to the PNP classification anyway if the analysis were to be extended to include pronominal possessors as well as nominal possessors (40). Whether the latter (**cls**) exists in the PNP domain is a question for further research.

(40) Finnish *ystävä-ni* [friend-1POSS] 'my friend'

Koptjevskaja-Tamm (2002; 2004) distinguishes two types of adnominal possessive: 'anchoring' and 'non-anchoring'. In an anchoring construction (41a), the noun in

the genitive case (the possessor) serves as an *anchor*, or reference point, for identifying (or grounding) the head, which is an individual (or set of individuals).

(41) Lithuanian (Koptjevskaja-Tamm 2004: 155–156)
a. *Petr-o namas* [Peter-GEN house] 'Peter's house'
b. *auks-o žedas* [gold-GEN ring] 'a gold ring'

In a non-anchoring construction (41b), on the other hand, the noun in the genitive serves as a modifier of the head. The constructions are otherwise identical and Koptjevskaja-Tamm treats them both as adnominal possessives: "The rationale for a similar treatment of anchoring and non-anchoring relations is obvious – both types of adnominal dependents characterize entities via their relations to other entities" (2004: 156). Paraphrasing Koptjevskaja-Tamm (2002: 154), Croft (2022) summarizes the features that differentiate non-anchoring from anchoring constructions as follows:

1. the object modifier is only type identifiable;
2. the modifier-head combination refers to a subclass of a broader class and often functions as a classificatory label for it, suggesting that the modifier and the head together correspond to one concept, but
3. the head cannot be identified via its relation to the modifier.

Croft also generalizes from adnominal possessive to 'nominal modifier construction', a term which covers any construction in which an object is used for modification.[23] The term thus applies to one of the cells in the Scapa Grid of basic cross-linguistic constructions (see Table 2 on page 31).

From this it is clear that binominals, as defined in the present study, are essentially non-anchoring (or 'typifying')[24] nominal modifier constructions, albeit ones in which a diachronic process of lexicalization has proceeded to the point where the binominal is coming to be a unitary lexeme; binominals are basically lexicalized typifying constructions that represent the penultimate stage in a continuum

23 Note that in Croft's terminology, the term 'modification' is used only for modification of an object, never for "modification" of a property or action. The three types of modification are therefore object modification (a.k.a. nominal modification), property modification, and action modification.

24 Pepper (2020) and Croft (2022) prefer the term 'typifying' to 'non-anchoring', on the grounds that it is better to describe something in terms of what it does, rather than doesn't do: the key thing about typifying constructions is that they denote *types* (or classes, or more precisely subclasses) rather than *individuals*, as is the case with anchoring constructions, cf. *women's magazine* vs. *Peter's magazine*.

of constructions from anchoring object modification to reference, which Croft (2022) calls the 'modification-reference continuum (42).

(42)

A		B		C		D
anchoring construction	⟶	**typifying construction**	⟶	**binominal lexeme construction**	⟶	**simple noun**
modification.....						..... *reference*

The transition from stage C (binominal form) to stage D (simplex form) is relatively well documented and understood. It was discussed briefly above (§4.1.1) and exemplified with Eng. *lord* and Ger. *Messer* at the endpoint of the continuum, and Eng. *nostril* midway between stages C and D.

The difference between stage B (typifying construction) and stage C is largely based on the degree of lexicalization, which is hard to measure. (One way to do so would be to rely on a comprehensive dictionary of the language and regard lemmas as lexical (hence, binominal) and other typifying constructions as syntactic.

As for the transition from stage A (anchoring construction) to stage B (typifying construction), Koptjevskaja-Tamm (2004) provides a lot of qualitative evidence from European languages. Five cases are distinguished, one of them with two subtypes, as follows:

1. *Identical structures*. The same morphosyntactic strategy is used for both anchoring and typifying constructions. Exemplified by Lithuanian, Georgian, Daghestanian, Russian and Finnish, which use a genitive modifier and inflect nouns for case but lack articles and a grammaticalized definiteness-indefiniteness opposition.
2. *Similar structures*. The same morphosyntactic strategy is used for both anchoring and typifying constructions but articles (markers of definiteness) on the modifier are permitted with the former but not the latter. Exemplified by Italian (prepositional strategy) and Scottish Gaelic (genitival strategy).
3. *Differing morphological complexity*. Typifying constructions are morphologically less complex and/or looser than anchoring constructions Exemplified by (a) Albanian, Rumanian, Turkish and Kirmandji (dependent-marking) and (b) Mordvin and Armenian (head-marking).
4. *Loss of nominal autonomy: compounding*. The relational or indexical marker found in anchoring constructions is lost in typifying constructions, leading to a compound or juxtaposition strategy for the latter. Exemplified by Erzya (Mordvin) and Swedish.

5. *Loss of nominal autonomy: relational adjectives.* Typifying constructions use a derived, adjectival form of the modifying noun instead of case markers (i.e. case affixes or adpositions). Exemplified by Russian.

Koptjevskaja-Tamm's analysis is qualitative. In the next section we discuss how to turn this into a quantitative analysis and, in doing so, present a novel methodology for comparing non-binary constructions. In this way, we arrive at two Greenbergian universals concerning the relationship between adnominal possessives and binominals.

4.2.2 Comparing non-binary typologies

One of the main goals of typology is the discovery of universals, in particular implicational universals. Croft (2003: 53) illustrates the idea with an example drawn from Hawkins (1983: 84), his Universal XI′ "If a language has noun before demonstrative, then it has noun before relative clause". This implicational universal covers the following four logically possible types: (i) demonstrative and relative clause both follow the noun (NRel, NDem); (ii) relative clause precedes the noun and demonstrative follows the noun (RelN, NDem); (iii) relative clause follows the noun and demonstrative precedes the noun (NRel, DemN); and (iv) demonstrative and relative clause both precede the noun (RelN, DemN).

The implicational universal restricts language variation to types (i), (iii) and (iv), and excludes type (ii), and can be expressed in the form of a tetrachoric table (Table 5).

Table 5: Tetrachoric table for N+Dem and N+Rel.

	DemN	**NDem**
RelN	✓	–
NRel	✓	✓

This approach works fine when comparing two binary constructions. i.e., when there are two parameters, each with two possible values, resulting in four logically possible language types. This is the case with the demonstrative modifier construction and the relative clause construction. But if we want to compare binominal constructions with anchoring or typifying constructions, it becomes unmanageable, since Koptjevskaja-Tamm's PNP typology consists of eight types (the six in Figure 1, plus possessive compounds and relational adjectives; see §3.1), and

the binominal typology consists of nine. Representing this as an 8x9 table would clearly not be very helpful. Moreover, many languages employ multiple strategies in order to represent attributive possession, just as they do to express binominals (see the discussion of Polish NN.GEN and NA_{REL} binominals in Cetnarowska, this volume). This complicates the comparison even more.

A different approach is thus required, one that Koptjevskaja-Tamm has already pioneered. Observe that her comparison table and its five primary categories do not focus on the *values* assigned to each language, but on a characterization of the *relationship* between each language's primary anchoring and typifying strategies: for each language the relationship is essentially described as "identical", "similar" or "differing", with the latter amenable to subcategorization such that it encompasses the two cases of "loss of autonomy" in addition to differing morphological complexity. In the following analysis we adopt Koptjevskaja-Tamm's categories, but with a minor adjustment to make them more amenable to statistical analysis.

The five categories listed above may be used as simple nominal variables. However, the adjectives used to describe the first three categories suggest a potential for representation as ordinal variables: *identical* → *similar* → *different*.[25] Pepper (2020: 274–275) therefore replaces Koptjevskaja-Tamm's three "orderable" categories with a more fine-grained system of five categories that express the *degree of similarity* between anchoring and binominal constructions. Not surprisingly, since this is about different grades of a property (similarity), a naming system based on adjectives is not very useful, so Pepper proposes one based on adverbs that qualify the adjective 'identical', the Pwav scale[26] (cf. the Likert scale) (43).

(43) always → mostly → sometimes → rarely → never

Clearly, these categories need to be defined more precisely for the task at hand, but before doing so there is second issue that needs to be addressed: that of mixed languages. As the example of Polish shows (19c, d), languages may have more than one binominal strategy available to them; some have as many as six (and some, like Polish, as many as nine if the order of constituents is taken into account); most have at least four; and only seven of the 106 languages in Pepper's database exhibit only one. The question thus arises which strategy to select

25 Levshina (2015: 17) uses the five-point Likert scale ('strongly disagree' – 'disagree' – 'neither agree nor disagree' – 'agree' – 'strongly agree') as an example of an ordinal variable and points out that "the categories thus differ in order, but we do not know yet by how much.".

26 Pepper's (2020: 275) name for this scale has been abandoned for one that is less immediately narcissistic.

for the comparison with anchoring constructions. Fortunately, almost every language shows a preference for one type of binominal or another, and 70 of the 106 in Pepper's database can be said to have a dominant type according to Dryer's (2013) criterion for dominance: that a value is either the only one possible or the one that is more frequently used. The comparison to follow is thus based on what Pepper terms the 'primary binominal strategy', defined as the type that occurs most frequently; languages that have no clear preference (Äiwoo, Galibi Carib and Selice Romani) are deemed to have no such strategy. In addition, Pepper employs the term 'secondary binominal strategy' for any non-primary strategy that is 'common' (defined as occurring in at least 10% of the data for any given language). Having defined these terms, the five grades of 'identicality' in (43) can be operationalized as shown in Table 6.

The definitions themselves are, of course, particular to the actual constructions that we are investigating, but the Pwav scale itself has universal validity and could provide an additional tool for typologists, alongside tetrachoric tables and semantic maps, for use when comparing non-binary typologies.

Table 6: The Pwav scale operationalized for the comparison of binominal and anchoring constructions.

grade	description
always	the primary binominal strategy is identical to the primary anchoring strategy and there are no secondary binominal strategies
mostly	the primary binominal strategy is identical to the primary anchoring strategy but there also are secondary binominal strategies
sometimes	a secondary binominal strategy is identical to the primary anchoring strategy, *or* the primary binominal strategy is identical to a secondary anchoring strategy
rarely	a secondary binominal strategy is identical to one of the secondary anchoring strategies
never	binominal strategies and anchoring strategies are quite different

4.2.3 Two universals of nominal modification

Before we proceed with the quantitative analysis of the data, it should be noted that the Pwav scale loses some of the qualitative detail in Koptjevskaja-Tamm's model (for example, differences in complexity and morphological tightness), but it is perfectly possible to add that back in, by subdividing the basic categories, and we do so here with "never" in order to capture grammaticalization. This category has been split in order to highlight cases where the principal binominal

strategy is a grammaticalized form of the principal anchoring strategy, following one of eight pathways that involve fusion or loss of a single marker (44). Finally, the rump "never" category contains languages in which no anchoring strategy is the same as any of the binominal strategies.

(44)

jxt	**gen**	**gen**	**con**	**prn**	**prn**	**dbl**	**dbl**
↓	↓	↓	↓	↓	↓	↓	↓
cmp	**cmp**	**jxt**	**cmp**	**jxt**	**cmp**	**con**	**gen**

Each pathway in (44) represents a single step: either fusion, as in the case of **jxt** → **cmp**, or loss of a single morpheme, as in **gen** → **cmp**. Other pathways are conceivable (e.g. **prp** → **gen**), but these are not attested in Pepper's data.

In all five categories except "never", binominal constructions can be said to recruit one of the anchoring strategies, sometimes across the board ("always"), sometimes to a lesser degree ("mostly", "sometimes", "rarely"), and sometimes in a more "grammaticalized" form. When the anchoring strategy is not recruited, more often than not, it is compounding (i.e. **jxt** or **cmp**) that fills the void.

Figure 5 plots the numbers for the six categories. We observe that all but 12 of the 105 languages for which data was available (almost 90%) recruit an anchoring strategy for use in the formation of binominals. Of the twelve languages in the sample that do not, the majority (58%) use a compounding strategy, either **jxt** or **cmp**.

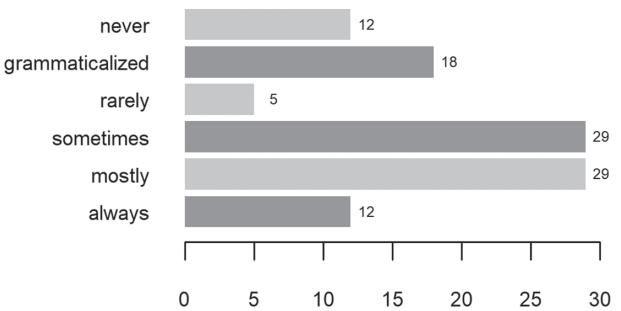

Figure 5: Grades of similarity across anchoring and binominal strategies.

Based on his analysis of the data contained in the binominals database, Pepper posits the following universals:

(45) *With overwhelmingly greater than chance frequency, languages recruit at least one of their binominal strategies from an anchoring nominal modifier construction.*

(46) *If a language does not recruit at least one of its binominal strategies from an anchoring nominal modifier construction, there is a strong tendency for it to use compounding for this purpose.*

5 Conclusion

In this chapter I offered four different definitions of binominal lexeme to complement the informal definition as "noun-noun compounds and their functional equivalents". Although couched in a variety of terms based on different theoretical frameworks, these have essentially identical extensions.

I then presented a nine-way classification of binominals: **jxt**, **cmp**, **der**, **cls**; **prp**, **gen**, **adj**, **con**; and **dbl**. These are represented on a two-dimensional grid that captures the number of markers, the locus of marking and the degree of fusion. The grid reveals two lacunae or "missing types": **prn** and **nml**. Whereas the first of these probably exists somewhere in the world's languages, the second seems to be a logical impossibility. The chapter also discussed intermediate types and various grammaticalization pathways.

Finally, I examined the relationship between binominal constructions and anchoring nominal modifier constructions and introduced a new methodology, based on the "Pwav scale", for comparing two non-binary constructions. This resulted in two Greenbergian universals concerning the recruitment of binominal strategies from nominal modifier strategies.

References

Adelaar, Alexander. 2009. Malagasy vocabulary. In Martin Haspelmath & Uri Tadmor (eds.), *World Loanword Database*. Leipzig: Max Planck Institute for Evolutionary Anthropology. http://wold.clld.org/vocabulary/28.

Aikhenvald, Alexandra Y. 2000. *Classifiers: A typology of noun categorization devices*. Oxford: Oxford University Press.

Aikhenvald, Alexandra Y. 2017. A typology of noun categorization devices. In Alexandra Y. Aikhenvald & R. M. W. Dixon (eds.), *The Cambridge handbook of linguistic typology*, 361–404. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781316135716.

Allan, Keith. 1977. Classifiers. *Language* 53(2). 285–311.
Bauer, Laurie. 2000. Word. In Geert E. Booij, Christian Lehmann & Joachim Mugdan (eds.), *Morphologie: ein internationales Handbuch zur Flexion und Wortbildung*, vol. 1, 247–257. Berlin: Mouton de Gruyter.
Bauer, Laurie. 2001. Compounding. In Martin Haspelmath, Ekkehard König, Wolfgang Oesterreicher & Wolfgang Raible (eds.), *Language typology and language universals: An international handbook*, 695–707. Berlin: Mouton de Gruyter.
Bauer, Laurie. 2005. The borderline between derivation and compounding. In Wolfgang U. Dressler, Dieter Kastovsky, Oskar E. Pfeiffer & Franz Rainer (eds.), *Morphology and its demarcations: Selected papers from the 11th Morphology meeting, Vienna, February 2004*, 97–108. Amsterdam: John Benjamins.
Bauer, Laurie & Elizaveta Tarasova. 2013. The meaning link in nominal compounds. *SKASE Journal of Theoretical Linguistics* 10(3). 2–18.
Bauer, Winifred. 1993. *Maori*. London: Routledge.
Booij, Geert. 2010. *Construction Morphology*. Oxford: Oxford University Press.
Bybee, Joan L. 1985. *Morphology: A study of the relation between meaning and form*. Amsterdam: John Benjamins.
Corbett, Greville G. 1987. The morphology/syntax interface: Evidence from possessive adjectives in Slavonic. *Language* 63(2). 299–345.
Creissels, Denis. 2009. The construct form of nouns in African languages: A typological approach. In Peter Austin, Oliver Bond, Monik Charette, David Nathan & Peter Sells (eds.), *Proceedings of Conference on Language Documentation & Linguistic Theory 2*, 73–82. London: SOAS.
Creissels, Denis. 2017. Construct forms of nouns in typological perspective. Unpublished paper, SLE 2017. http://www.deniscreissels.fr/public/Creissels-Cstr.pdf.
Croft, William. 1990. *Typology and universals*. Cambridge: Cambridge University Press.
Croft, William. 1991. *Syntactic categories and grammatical relations: The cognitive organization of information*. Chicago: The University of Chicago Press.
Croft, William. 2001. *Radical Construction Grammar: Syntactic theory in typological perspective*. Oxford: Oxford University Press.
Croft, William. 2003. *Typology and universals*. 2nd ed. Cambridge: Cambridge University Press.
Croft, William. 2022. *Morphosyntax: Constructions of the world's languages*. Cambridge: Cambridge University Press.
Dixon, R. M. W. 2010. *Basic linguistic theory. Volume 2: Grammatical topics*. Oxford: Oxford University Press.
Dokulil, Miloš. 1962. *Tvoření slov v češtině. 1: Teorie odvozování slov*. Praha: Československé akademie věd.
Dokulil, Miloš. 1966. Zum wechselseitigen Verhältnis zwischen Wortbildung und Syntax. *Travaux linguistiques de Prague* 1. 215–224.
Dokulil, Miloš. 1994. The Prague School's theoretical and methodological contribution to "word-formation" (derivology). In Philip A. Luelsdorff (ed.), *The Prague School of structural and functional linguistics: A short introduction*, 123–161. Amsterdam: John Benjamins.
Dryer, Matthew S. 2013. Determining dominant word order. In Matthew S. Dryer & Martin Haspelmath (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. https://wals.info/chapter/s6.

Floricic, Franck. 2016. French. In Peter O. Müller, Ingeborg Ohnheiser, Susan Olsen & Franz Rainer (eds.), *Word-formation: An international handbook of the languages of Europe. Vol. 4*, 2661–2682. Berlin: Mouton de Gruyter.

Haspelmath, Martin. 1996. Word-class-changing inflection and morphological theory. In *Yearbook of Morphology 1995*, 43–66. Springer. http://link.springer.com/chapter/10.1007/978-94-017-3716-6_3.

Haspelmath, Martin. 2009. Terminology of case. In Andrej Malchukov and Andrew Spencer (eds.), *The Oxford handbook of case,* 505–517. Oxford: Oxford University Press.

Haspelmath, Martin. 2010. Comparative concepts and descriptive categories in crosslinguistic studies. *Language* 86(3). 663–687.

Haspelmath, Martin. 2011. The indeterminacy of word segmentation and the nature of morphology and syntax. *Folia Linguistica* 45(1). 31–80. https://doi.org/10.1515/flin.2011.002.

Haspelmath, Martin. 2012. How to compare major word-classes across the world's languages. *UCLA Working Papers in Linguistics, Theories of Everything* 17, Article 16. 109–130.

Haspelmath, Martin. 2020. The morph as a minimal linguistic form. *Morphology* 30(2). 117–134. https://doi.org/10.1007/s11525-020-09355-5.

Hawkins, John A. 1983. *Word order universals*. New York: Academic Press.

Khalilov, Madzhid & Zaira Khalilova. 2016. Bezhta. In Peter O. Müller, Ingeborg Ohnheiser, Susan Olsen & Franz Rainer (eds.), *Word-formation: An international handbook of the languages of Europe. Vol. 5*, 3658–3678. Berlin: Mouton de Gruyter.

Kiefer, Ferenc. 2009. Uralic, Finno-Ugric: Hungarian. In Rochelle Lieber & Pavol Štekauer (eds.), *The Oxford handbook of compounding*, 527–541. Oxford: Oxford University Press.

Koptjevskaja-Tamm, Maria. 2002. Adnominal possession in the European languages: Form and function. *STUF – Language Typology and Universals* 55(2). 141–172.

Koptjevskaja-Tamm, Maria. 2003. Possessive noun phrases in the languages of Europe. In Frans Plank (ed.), *Noun phrase structure in the languages of Europe*, 621–722. Berlin: Mouton de Gruyter.

Koptjevskaja-Tamm, Maria. 2004. Maria's ring of gold: Adnominal possession and non-anchoring relations in the European languages. In Kim Ji-yung, Yury A. Lander & Barbara H. Partee (eds.), *Possessives and beyond: Semantics and syntax*, 155–181. Amherst, MA: GLSA Publications.

Langacker, Ronald W. 2008. *Cognitive Grammar: A basic introduction*. Oxford: Oxford University Press.

Levshina, Natalia. 2015. *How to do linguistics with R: Data exploration and statistical analysis*. Amsterdam: John Benjamins.

Müller, Peter O., Ingeborg Ohnheiser, Susan Olsen & Franz Rainer (eds.). 2015. *Word-formation: An international handbook of the languages of Europe. 5 vols*. Berlin: de Gruyter Mouton.

Nagórko, Alicja. 2016. Polish. In Peter O. Müller, Ingeborg Ohnheiser, Susan Olsen & Franz Rainer (eds.), *Word-formation: An international handbook of the languages of Europe. Vol. 4*, 2831–2852. Berlin: de Gruyter Mouton.

Newman, Paul. 2000. *The Hausa language: An encyclopedic reference grammar*. New Haven: Yale University Press.

Nichols, Johanna. 1992. *Linguistic diversity in space and time*. Chicago: The University of Chicago Press.

Pepper, Steve. 2010. *Nominal compounding in Nizaa: A cognitive perspective*. SOAS University of London Master's thesis.

Pepper, Steve. 2016. Windmills, Nizaa and the typology of binominal compounds. In Lívia Körtvélyessy, Pavol Štekauer & Salvador Valera (eds.), *Word-formation across languages,* 281–310. Newcastle: Cambridge Scholars Publishing.
Pepper, Steve. 2018. Onomasiological types and the typology of binominal lexemes. Paper presented at the *Typology and Universals in Word-Formation IV* conference, Košice 2018. https://www.academia.edu/38060850.
Pepper, Steve. 2020. *The typology and semantics of binominal lexemes: Noun-noun compounds and their functional equivalents*. Oslo: University of Oslo PhD dissertation. https://www.academia.edu/42935602.
Pepper, Steve. in prep. Onomasiological types and the typology of complex nominals.
Ralli, Angela. 2013. *Compounding in Modern Greek*. Dordrecht: Springer.
Russell, Joan. 2003. *Swahili* (Teach Yourself Books). London: Hodder and Stoughton.
Štekauer, Pavol. 1998. *An onomasiological theory of English word-formation*. Amsterdam: John Benjamins.
Štekauer, Pavol. 2005. Onomasiological approach to word-formation. In Pavol Štekauer & Rochelle Lieber (eds.), *Handbook of word-formation*, 207–232. Dordrecht: Springer.
Tuggy, David. 1992. The affix-stem distinction: A Cognitive Grammar analysis of data from Orizaba Nahuatl. *Cognitive Linguistics* 3(3). 237–300.
Uluhanov, Igor' S. 2016. Russian. In Peter O. Müller, Ingeborg Ohnheiser, Susan Olsen & Franz Rainer (eds.), *Word-formation: An international handbook of the languages of Europe, Vol. 4*, 2953–2978. Berlin: Mouton de Gruyter.
Urban, Matthias. 2012. *Analyzability and semantic associations in referring expressions: A study in comparative lexicology*. Leiden University PhD dissertation.
Wälchli, Bernhard. 2005. *Co-compounds and natural coordination*. Oxford: Oxford University Press.
Watson, Rachel. 2015. *Kujireray: morphosyntax, noun classification and verbal nouns*. SOAS University of London PhD dissertation.
Wilkens, Jens. 2016. Turkish. In Peter O. Müller, Ingeborg Ohnheiser, Susan Olsen & Franz Rainer (eds.), *Word-formation: An international handbook of the languages of Europe. Vol. 5*, 3367–3385. Berlin: de Gruyter Mouton.
Yeon, Jaehoon & Lucien Brown. 2011. *Korean: A comprehensive grammar*. London: Routledge.