Francesca Masini, Simone Mattiola and Steve Pepper
Exploring complex lexemes cross-linguistically
Editors' introduction

1 Towards a typology of complex lexemes

Concept-naming is one of the most fundamental activities performed by speakers, who need either ready-made labels to talk about entities or devices to build new labels (be they rules or processes, schemas or analogical mechanisms). Knowing how languages perform the basic function of creating labels to name concepts, especially complex concepts, is crucial to understanding their creative potential in building new (potentially stable) categories and, more generally, to understanding how they (may) categorize reality, and refer to it. What are the strategies employed by languages for naming complex concepts? How do they differ cross-linguistically, and what are the limits of their variation? Are there strategies that are more widespread than others, or even universal?[1]

These are questions for lexical typology and/or word-formation typology, but what we know about the typology of complex concept naming is very limited compared to what we know about domains like word order or inflectional morphology. There may be different reasons behind this state-of-affairs. Analysing all of them falls outside the scope of the present introduction: we will just discuss some factors that we deem relevant for our current purposes.

Complex concept naming is definitely related to word-formation. The domain of word-formation can count on an extremely rich and ever-growing body of literature, which would be impossible to credit here (suffice it to mention the collections edited by Booij, Lehmann & Mugdan 2000; 2004; Lieber & Štekauer 2009a; 2014; Müller et al. 2015; Lieber et al. 2021). However, quite surprisingly, word-formation has rarely been the subject of large-scale, thorough typological

---

1 We wish to thank the many anonymous referees who generously agreed to review the chapters included in this volume, including this introduction: their insightful comments significantly improved the quality of the volume. Heartfelt thanks are also due to Jean-Christophe Verstraete for his guidance and constant support, which were essential to bring this project to conclusion. We are also grateful to the audience of the *When "noun" meets "noun"* workshop at the 50th Annual Meeting of the Societas Linguistica Europaea (Zurich, 10–13 September 2017). The usual disclaimers apply.

investigations (beyond European languages, cf. Müller et al. 2015–2016) despite some laudable initial attempts (see Štekauer, Valera & Körtvélyessy 2012), with the result that gaining comparable data on word-formation processes is not an easy task. Possibly as a consequence of this lack of typological studies, word-formation is under-represented in major typological online resources. Take, for instance, WALS Online (Dryer & Haspelmath 2013), where few features (13 out of 192) pertain to the "Lexicon" and where the "Morphology" area is entirely devoted to inflection, word-formation being basically absent, with the notable exception of reduplication (Rubino 2013). A similar picture emerges consulting APiCS Online (Michaelis et al. 2013).

On the other hand, the rise of lexical typology in the last couple of decades (Koch 2001; Brown 2001; Koptjevskaja-Tamm 2008; Koptjevskaja-Tamm, Rakhilina & Vanhove 2015) has contributed to bringing the lexicon back to typologists' attention. So far, however, these studies have mostly focused – quite understandably – on (simple) words and on lexical semantics (e.g. Vanhove 2008), rather than complex words and the devices that create them. Still, the latter issue falls within the scope of the field and would definitely deserve to be explored more fully, as Koptjevskaja-Tamm & Veselinova (2020: §2.2.3) have recently argued: "word formation is the research domain where the overlap between the lexical typology and morphology is particularly salient. However, systematic cross-linguistic research on word-formation strategies and their functions has so far been modest. This is all the more surprising given the abundance of data on word formation in individual languages and in individual language families". The authors advocate for a deeper interaction between the fields of morphology and lexical typology ("a huge, still very much unexplored domain"), and report, more generally, that "there is still no methodology that paves the way for a systematic comparison of the vocabularies of different languages".

Still another issue that arises for complex concept naming is the array of formal strategies that may perform this function, the different status these strategies may have in different research traditions, and the frequent lack of connection in the literature between different strategies. In morphology, the main mechanisms for creating complex words are derivation and compounding (Lieber & Štekauer 2009a; 2014; Štekauer, Valera & Körtvélyessy 2012), to which a variety of other mechanisms can be added, from reduplication to incorporation, from conversion to subtraction, from blending to clipping, and so on. Of course, simple words may convey a complex concept, too. And some types of multi-word expressions (cf. Baldwin & Kim 2010; Hüning & Schlücker 2015), like 'phrasal lexemes' (Booij 2009; Masini 2009), also perform a clearly concept-naming function. However, they are generally not considered as part of morphology, being objects beyond the word level (whatever the boundary may be).

This separation is a reflection of the traditional divide between morphology and syntax that has characterized linguistics since structuralism. The wealth of literature on the word/phrase (and compound/phrase, e.g. Lieber & Štekauer 2009b) distinction is symptomatic in this respect. However, when your goal is to understand how complex concept naming works, it is quite clear that *all* these strategies should be considered and kept together, for the simple reason that, potentially, *all* of them may express complex concepts. Suffice it to consider the crosslinguistic variation we encounter, in terms of morphosyntactic strategies, when we translate terms from one language to another. A trivial illustration follows for 'earthquake' in a few (related and unrelated) languages:[2]

(1) a. English [ENG]
         earthquake   NN compound
     b. Italian [ITA]
         sisma   simple N
     c. French [FRA]
         tremblement de terre (quake of earth)   NPrepN phrasal lexeme
     d. Polish [POL]
         trzęsienie ziemi (quake earth.GEN)   NN$_{GEN}$ phrasal lexeme
     e   Hebrew [HEB]
         רעידת אדמה [*re'idat adama*] (quake.CS earth) construct state
     f. Arabic [ARA]
         زلزال [*zalzāl*]   (non-concatenative) derivation

Different strategies may surface in one and the same language (cf. French *tremblement de terre* vs. *séisme*, a simple word), possibly competing with one another, making the whole picture even more complicated.

Masini (2019a) discusses the relationship between word-formation and multi-word expressions and makes the case for a unified approach to these domains – both in theoretical linguistics and in typology – by virtue of their common function. She also advocates for a more active inclusion of multi-word expressions within

2 List of abbreviations used in this chapter: ADJZ = adjectivizer; CLF = classifier; CON = connective; CS = construct state; GEN = genitive; NMLZ = nominalizer; OBL = oblique; PREP = preposition; REP = reported. As for language classification, we decided not to include genealogical information for the languages cited in this introduction, and in the whole volume, since there is no general agreement in the linguistic community about genealogical classification and language names. Rather, we decided to provide the ISO-code 639-3 (or the Glottocode, where no ISO-code exists) for each language when first introduced.

lexical typology, as also argued by Koptjevskaja-Tamm & Veselinova (2020). Taking into greater consideration word-formation processes and multi-word expressions is paramount to get a more truthful and comprehensive picture of vocabularies across languages. Indeed, both complex words and at least a section of multi-word expressions, viz. phrasal lexemes, are employed to create new complex lexical items, sometimes co-existing within one and the same language, where they may compete with each other, or compensate for one another in the expression of lexico-conceptual meanings (Masini 2009b).

The success of a 'unified' typological approach to complex lexemes depends on two conditions: (i) a supporting theoretical background; and (ii) clear definitions for cross-linguistic comparison.

As for (i), we believe that Construction Grammar approaches (Goldberg 1995; 2006; 2019; Hoffmann & Trousdale 2013; Hilpert 2014) may be the answer, for at least two reasons.

On the one hand, Construction Grammar has proved to be well-equipped to deal with cross-linguistic explorations, both contrastive studies (e.g. Hilpert & Östman 2014) and typology (Croft 2001). Crucially, regarding constructions (i.e., form-function pairings) as the basic units of linguistic analysis overcomes the traditional subdivision into levels of analysis, especially between syntax and the lexicon, thus accommodating quite straightforwardly a wide array of 'intermediate' structures like idioms and multi-word expressions (cf. the syntax-lexicon continuum as depicted by Croft 2001: 17; Goldberg 2013: 17). Studies within Construction Morphology (Booij 2010; Masini & Audring 2019) are also explicit about including multi-word expressions into the picture, allowing for a unified treatment of 'lexemes', namely constructions with a concept-naming function, independently of their structural complexity and internal composition (simple words, complex words, multi-word units, etc.). In short, a constructionist view 'licenses' an onomasiological approach to lexemes in general, and more specifically to 'complex lexemes', namely non-basic lexemes which reflect speakers' creativity along with their need to label new categories using previously existing, meaningful linguistic material (much in the vein of Martinet's 1949 'primary articulation').

Incidentally, since constructions may have different levels of schematicity or abstractness, we can take into consideration both lexically specified constructions (for instance, existing complex lexemes like *earthquake*) and more abstract constructions (for instance, NN compounding) that function as templates for the creation of new complex lexemes, since both are regarded as constructions or 'signs'. Assessing the productivity of lexical constructions may not be easy (especially in large-scale typological studies, which often must rely on data elicitation or descriptive grammars). However, despite the practical challenges it poses, productivity may be an important factor when it comes to understanding the inter-

play between the strategies actually available in a language (family) to convey new lexico-conceptual meanings.

As for (ii), the comparability of linguistic structures is one of the most often debated questions in linguistic typology. Although this is a well-known problem dating back at least to the first modern descriptions of languages of Native North America (Boas 1911), the last couple of decades have witnessed an intense debate in the typological literature. Starting from Dryer (1997) and Croft (2001), and more recently Haspelmath (2007; 2010) and Cristofaro (2009), grammatical categories have come to be conceived as specific entities of single languages (*language-specific*) or even of specific constructions (*construction-specific*). It is by now widely assumed that cross-linguistic comparison should not be based on pre-established linguistic categories: "The most important consequence of the non-existence of pre-established categories for language typology is that cross-linguistic comparison cannot be category-based, but must be substance-based, because substance (unlike categories) is universal" (Haspelmath 2007: 124). Instead, cross-linguistic comparison should be conducted via comparative concepts, namely "concepts created by comparative linguists for the specific purpose of crosslinguistic comparison. Unlike descriptive categories, they are not part of particular language systems and are not needed by descriptive linguists or speakers. They are not psychologically real, and they cannot be right or wrong. They can only be more or less well suited to the task of permitting crosslinguistic comparison. They are often labeled in the same way as descriptive categories, but they stand in a many-to-many relationship with them [. . .]. Comparative concepts are universally applicable, and they are defined on the basis of other universally applicable concepts: universal conceptual-semantic concepts, general formal concepts, and other comparative concepts" (Haspelmath 2010: 665).

Applying this to the problem of complex lexemes, we are by now aware that 'word' is a cross-linguistically unreliable and tricky concept (cf. among many others, Ramat 1990; Haspelmath 2011; Arkadiev & Klamer 2019). But, most importantly, it is the wrong concept for the goal we are pursuing here, namely a typology of 'complex lexemes' intended as non-basic lexemes which also include objects that are *not* morphological 'words' (definable according to language-specific criteria).

What we would like to propose at this point is to consider 'complex lexeme' as a comparative concept. A first (admittedly tentative) definition of this category could be the following:

(2) A complex lexeme (CL) is a concept-naming unit, with a (potentially) stable denotation in a language, which combines at least two formatives or is the result of a (non-concatenative) formal operation over a formative, and which combines at least two concepts entertaining some semantic relation.

This definition appears broad enough to encompass complex words of various types, multi-word expressions and possibly other naming units a language may display that do not fall within these two classes, leaving their formal demarcation and definition to language-specific criteria which are simply not relevant for their inclusion into the CL category. However, this definition also poses some questions. Some clarifications are therefore in order.

One concerns the word 'potentially'. According to the definition in (2), a CL should have a concept-naming, labelling function, which is at least *potentially* stable in the system: nonce expressions, formed on the spur of the moment according to some template of lexeme creation, are not stable by definition, but have the potential to become conventionalized signs, given the appropriate conditions. So the word 'potentially' merely serves to keep nonce formations into the picture.

Another relates with the word 'unit', which is used here to express that CLs should be endowed with some degree of internal cohesion, to be ascertained and defined according to language-specific criteria.

Still another issue is 'complexity'. According to the definition in (2), the complexity of CLs is twofold, regarding both its formal and its functional side. On the formal side, we include items which are either the combination of two or more formatives (affixes, words, clitics, classifiers, etc.) or the result of the application of some other type of operation which does not involve the addition of morphosyntactic material (think of clipping or conversion). On the functional side, our definition states that CLs should combine at least two concepts. This may be a weak point of the definition because it seems to imply an iconic relation between form and function which is far from real. As a matter of fact, the world's languages display both basic (formally simple) lexical items that 'conflate' (Talmy 2007) two or more concepts (see Italian *sisma* in (3)), and formally complex lexical items that convey basic notions, as pointed out by Koptjevskaja-Tamm & Veselinova (2020). These facts should obviously be taken into account when trying to answer the wider question, raised by lexical typologists, of which meanings can/cannot be conveyed by simple/basic items or by complex/non-basic items. However, we think it might be fruitful to restrict CLs to items that are both formally and conceptually complex, partly for sheer convenience, partly because this restriction might help to focus on the task at hand: what are the strategies employed by languages for naming complex concepts (we already know simple words are one of these but we are far from having a full and typologically-informed picture of everything else) and what are the principles behind their cross-linguistic variation and distribution.

Obviously, a far-reaching typology of CLs as defined above is much easier to conceive than to actually build. The domain to be covered is vast indeed. However, it looks more feasible if one constrains the domain of investigation to something more manageable in terms of coverage, by creating (ad hoc) 'daugh-

ter comparative concepts' from the more general 'complex lexeme' comparative concept proposed in (2). For instance, the definition in (2) may be constrained in terms of the kind of formatives one wants to focus on (affixes vs. words, prefixes vs. affixes), the kind of processes involved (concatenative vs. non-concatenative), the nature of concepts being combined, or a combination of these factors.

This is the choice we have made in this volume, by putting the spotlight on a specific kind of CL, namely 'binominal lexemes', which will be introduced and defined in Section 2 (and Chapter 2). With this 'case-study', we would like to make the case that, by delimiting the range of possible combinations of concepts and formatives, in either structural or functional terms, doing 'complex lexeme typology' becomes an enterprise wihin the bounds of possibility.

2 Focus on binominal lexemes

The act of naming a (new) complex concept through the combination of two existing concepts can be seen most clearly in determinative noun-noun compounds, such as those in (3), all of which denote the meaning 'railway'.

(3)   a. German [DEU] *Eisenbahn* [iron.way]
       b. Mandarin Chinese [CMN] 铁路 *tie3 lu4* [iron road]
       c. Mapudungun [ARN] *trenrüpü* [train.way]
       d. Saramaccan [SRM] *talán fútu* [train foot]

In each of these examples, the first constituent (the 'modifier') serves to restrict the extension of the class of objects denoted by the second (the 'head'): a railway is conceptualized as a kind of way (or road) that is somehow related to the concept IRON (3a)-(3b) or TRAIN (3c), or as a kind of foot that is related to the concept TRAIN (3d).[3] They are all instances of nominal modification constructions (Croft 2001; 2022).

Now, one of the most interesting aspects of such compounds is precisely the fact that the relationship between the two combining concepts is not stated explicitly: the motivation for combining the two concepts in question must be inferred by the user. In the case of (3a)-(3b), the choice of concepts is clearly motivated by a conceptualization of 'railway' as a way that is MADE OF iron; in

---

3 Naturally, these conceptualizations are particularly relevant during the coining and adoption of the new expression, and less so once it has become conventionalized.

(3c) a railway is conceptualized as a way that is USED BY trains; and in (3d) it is conceptualized metaphorically as a 'foot' that belongs to, or is PART OF a train.

Investigations into the range of semantic relations exhibited by determinative noun-noun compounds have the potential to reveal interesting insights into the associative nature of human thought (e.g., what are the most salient relationships, which ones are used for which kinds of complex concepts, what differences can be found across languages, etc.).

However, if *Eisenbahn* (3a) provides interesting evidence in this regard, so too do the complex lexemes in (4), all of which again denote the meaning 'railway'.

(4)   a.   French *chemin de fer* [track PREP iron]
        b.   Russian [RUS] железная дорога *železnaja doroga* [iron.ADJZ road]
        c.   Modern Hebrew מסילת ברזל *mesilat barzel* [track.CS iron]
        d.   Bezhta [KAP] *kilos hino* [iron.OBL.GEN way]

The four complex lexemes in (4) are all functionally equivalent to those in (3) in that they all combine items denoting two concepts to denote the new (complex) concept 'railway'; in fact, the same two concepts as in (3a) and (3b): IRON and WAY.

But are they compounds? The answer to that question varies with the linguistic tradition of each individual language:

–   For French, Floricic (2016) makes a distinction between compounds *stricto sensu* (or "compounds proper") and a subtype dubbed 'synapsie', a term which traces back to Benveniste (1966), "which is syntactic in essence and consists of a group of lexemes connected by a linker: *pomme de terre* 'potato; lit. apple of earth', *chemin de fer* 'railway; lit. way of iron', etc.".
–   Uluhanov's (2016) discussion of word-formation in Russian makes no mention at all of the use of relational adjectives (like *železnaja*) in the formation of complex lexemes, let alone including them under 'Composition'.
–   Levi's (1976) paper on Hebrew "compound nominals" of the type exemplified by (4c) generally avoids the term 'compound' itself, preferring the label traditionally used in Hebrew linguistics '*smixut* construction' (namely, construct state construction).
–   Finally, Khalilov & Khalilova's (2016) coverage of composition in Bezhta allows that compounds "can also be based on the oblique nominal stem combined with a noun in the genitive case", as is the case in (4d).

What this tells us is that 'compound' is not a suitable term for use in cross-linguistic comparison, and it is precisely for this reason that we introduce the comparative concept 'binominal lexeme'. Informally, a binominal lexeme is simply a *noun-*

noun compound or its functional equivalent. (See Pepper, this volume, a for formal definitions, further examples and a typology of binominal lexemes.)

The term binominal lexeme covers all the examples in (3) and (4), despite the fact that they exhibit a range of morphosyntactic strategies (compounding, juxtaposition, prepositional, adjectival, construct and genitival). Given the onomasiological definition briefly introduced here and further developed in Chapter 2, it also covers the examples in (5), which clearly are not compounds.

(5) a. Slovak [SLK] *železnica* [iron.ADJZ.NMLZ] 'railway'
     b. Murui Hitoto [HUU] *ui.tïraï* [eye.CLF(hair)] 'eyelash'[4]

(5a) is a denominal derivation that parallels the Russian example in (4b), except that it employs a nominalizing suffix (*-ica*) with the general meaning 'thing' instead of a head noun meaning 'road' (*doroga*). (5b) is an example of a classifier construction in which the classifier *tïraï* has a derivational rather than a classificatory function. The former embodies the MADE OF relation, like (3a), (3b) and (4); and the latter embodies the PART OF relation, like (3d).

With this new comparative concept it becomes possible to study one of the most important types of complex lexeme formation found in the world's languages: that in which two object (or "nominal") concepts are combined to denote a new meaning.

The contents of the present volume show that binominal lexemes can be studied from a variety of perspectives:

– **Morphosyntactic strategies.** The formal mechanisms employed in creating binominals can be investigated:
    – cross-linguistically and in their totality, with the aim of exploring a particular typological classification (see Pepper 2020; this volume, a);
    – with a cross-linguistic focus on one particular strategy (see Creissels, this volume);
    – in their totality within a language family (see Lesage, this volume) or within a particular language (see van Egmond, this volume; Næss, this volume);
    – via a selection of strategies in a particular language (see Cetnarowska, this volume);

---

4 Not surprisingly, speakers of the Amazonian language Murui Hitoto do not have a word for 'railway'. If they needed one, it might conceivably be *yoezo* [metal.CLF.REP:path], parallel to *yoeya* [metal.CLF:craft] 'boat made of iron, metal', with the repeater *-zo* 'path' used instead of the classifier *-ya* 'craft', cf. Wojtylak (2017: 194).

- – contrastively, between unrelated languages (see Rose and Van linden, this volume);
  - – within a particular semantic domain (see Naccarato and Huang, this volume).

- **Semantic relations.** Again, these can be investigated cross-linguistically, for example, with a view to developing a unified classification (see Pepper 2020; this volume, b); within a language family (see Károly, this volume) or a particular language (see Gürer, this volume); or contrastively (see Pleshak, this volume).

- **Language acquisition.** The acquisition of binominals and competition between different binominal strategies can be investigated within individual languages or cross-linguistically (see Rosenberg, this volume).

These contributions, each of which could serve as a model for further work, will be introduced in more detail in the next section. The point to be made here is the diversity of research questions that they embody. However, there are many other potential paths of investigation that are not represented in the present volume. It will suffice here to mention two:

- **Correlations between form and meaning.** It might be expected that there is some kind of correlation between morphosyntactic strategies and the semantic relations found in binominals. According to Pepper (2020: §7.3) this is not the case from a typological perspective. However, such correlations are likely to be found in many specific languages, as first demonstrated for Nizaa (SGI) by Pepper (2010). This is clearly an area that needs more research (see Pepper 2020: §7.3.1 for a list of 22 languages from his sample of 106 that he finds worthy of further research in this respect).

- **Lexico-constructional patterns**: The choice of which two concepts to combine in order to denote a complex concept – in other words, to create a complex lexeme – varies across languages, and is influenced by both language contact, cultural considerations and language-internal resources. As an example, Figure 1 shows how concepts are combined to denote 'railway' in a sample of 57 languages (Pepper 2018a).

In short, the comparative concept of binominal lexeme – a suitably restricted and more manageable 'daughter comparative concept' of the more general 'complex lexeme' – offers rich scope for investigations into the nature of the (complex) lexicon. It can also serve as a model for further 'sibling' comparative concepts, such as those that involve property modification or action modification rather than object modificaton.
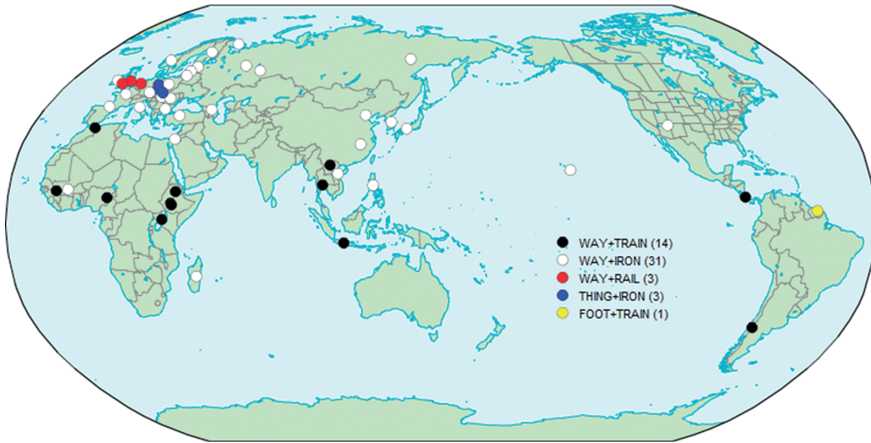
Figure 1: Lexico-constructional patterns for 'railway'.

3 What this volume is about

The present volume is divided into three parts: the first consists of typological, contrastive and descriptive studies that focus primarily on morphosyntactic strategies; the second on studies that focus primarily on semantic relations; and the third on acquisition. There is overlap, of course – several papers in the first part include a discussion of semantics, and those in the second make reference to morphosyntactic strategies – but the groupings show a certain coherence.

Following this introduction, the volume itself opens with the first of two contributions by **Steve Pepper**, entitled *Defining and typologizing binominal lexemes*. The goal of this chapter is "to provide a general introduction to binominals", and he starts by offering four different definitions of binominal lexeme. Although couched in a variety of terms, based on different theoretical frameworks, these have essentially identical extensions. He then presents a nine-way classification of binominals based on the morphosyntactic strategies employed in the world's languages. These nine 'binominal types' are arranged on a 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 strategies that are theoretically possible, but so far unattested. After discussing these "missing types", the paper turns to the question of grammaticalization and examines both gradient binominal phenomena and the relationship between binominals and possessive constructions. In order to evaluate the latter relationship quantitatively, he

describes an innovate method for comparing two non-binary constructions, and he ends by stating two Greenbergian universals regarding binominals.

The chapter by **Denis Creissels**, entitled *Binominals and construct marking*, provides a cross-linguistic examination of construct marking as one particular type of binominal lexeme construction (labelled **con**[5] in Pepper's classification). Creissels starts by defining construct marking as a particular technique of marking relationships between head nouns and their dependents. He shows that nominal modification constructions involving construct marking can be found well beyond the language families (Semitic, Nilotic, and Oceanic) in which the term 'construct' is traditionally used, and illustrates the cross-linguistic variation. He then examines the relationship with binominal formation and shows that in languages that make use of construct marking in adpossessive constructions, it is common for construct markers to be used more or less productively in the formation of binominals.

In his contribution, *Compounds and other nominal modifier constructions in Pama-Nyungan languages*, **Jakob Lesage** develops his own set of comparative concepts that differs in two ways from those used in the rest of this volume. Firstly, he extends the concept of nominal modification to include (some) property modification constructions, on the grounds that "it is not possible to clearly distinguish nouns and adjectives in various languages of [his] sample and, where this may be possible, not all grammars provide enough data or analysis to make such a distinction". Secondly, his subdivision is made along functional lines rather than in terms of morphosyntactic strategies: he distinguishes between 'binominal compounds' (where the modifier has a classifying function rather than a qualifying function); 'descriptive phrases' (where the modifier has a qualifying function); 'generic-specific constructions' (combinations of a generic noun and a specific noun, between which there is a relation of hyponymy); and 'inalienable (attributive) possession constructions' (where the relation is meronymic).

Marie-Elaine van Egmond's chapter, *New types of binominal lexeme in Anindilyakwa (Australia)*, provides an intriguing description of four types of binominal lexeme in the polysynthetic Gunwinyguan language Anindilyakwa. The constructions involved are two possession constructions, one expressing inalienable possession and one indicating alienable possession, plus a proprietive suffix which has a 'having, being equipped with' meaning, and a privative construction that contributes a meaning of 'without'. In Van Egmond's analysis

---

5 The mnemonics for the nine types in Pepper's classification are **jxt**, **cmp**, **der**, **cls**; **prp**, **gen**, **adj**, **con**; and **dbl**. They are indicative of the morphosyntactic strategies employed but should not be taken literally (see Pepper, this volume, a).

three of these constructions are regarded as type **adj** in Pepper's classification, while the fourth (inalienable possession) is claimed to be an instance of Pepper's "missing type" **nml**. This analysis differs from that in Pepper (2020: 474) who classifies them as **gen** and **con**, respectively. The difference hinges on whether the affixes in question are regarded as transpositional or not (see Pepper, this volume, a: §2.3.2 for further discussion). Be this as it may, there is no disputing van Egmond's conclusion, that the study of typologically lesser-known languages may shed new light on the typology of binominals, as access to new data broadens the scope of the typological generalizations that we can achieve.

Åshild Næss' chapter, *Binominals in Äiwoo: Compounds, possessive constructions, and transitional cases*, describes the different morphosyntactic strategies available to form binominal lexemes in the Oceanic language Äiwoo, and the semantic relations that characterise the different construction types. The Äiwoo data show examples of two or three distinct types in Pepper's classification. Næss shows that in all but unambiguous compounding constructions, possessor indexing plays an important role. It is found on the head in the direct possessive construction, on the possessive classifier in the indirect possessive construction, and on the relational prepositions. In most of the ambiguous or transitional cases, the key question is whether possessor indexing is present and where it is located. Thus, Äiwoo shows a complex picture that cannot be fully accounted for in Pepper's typology: relational morphemes such as possessive classifiers and relational prepositions may also carry possessive marking, and it is in these cases that problems arise in assigning the Äiwoo constructions to appropriate types.

Bożena Cetnarowska's chapter, *NN.ɢᴇɴ and NAʀᴇʟ juxtapositions in Polish: Syntactic schemas employed in building phrasal nouns*, analyzes some multiword unit structures of Polish. More specifically, the author focuses on two types of what is called 'juxtaposition' in the Polish grammatical tradition, that is, noun + genitive noun (NN.ɢᴇɴ) and noun + relational adjective (NAʀᴇʟ). Both of these structures are binominal lexemes according to the definition adopted in the present volume (**gen** and **adj** types respectively). Since these are considered as phrasal units, Cetnarowska also compares them with another competing binominal, morphological compounds (or *compound proper*) (**cmp**). The main difference that emerges lies in their semantics: while **gen** and **adj** are endocentric, **cmp** is exocentric.

The chapter by **Françoise Rose** and **An Van Linden** (*The derivational use of classifiers in Western Amazonia*) investigates the distinction between two different types of binominal lexeme, that is, classifier derived nouns (**cls**) and noun-noun compounds (**cmp**), in two Western Amazonian languages, Mojeño Trinitario and Harakmbut, that are not genetically related or in contact with each other. This topic is particularly relevant since in several South American languages classi-

fiers and nouns are not always easily distinguishable. For example, in Mojeño Trinitario the two categories share the same syntactic distribution but are formally different, while in Harakmbut they are formally identical but have different syntactic distributions. Having analysed **cls** and **cmp** types in the two languages, the two authors propose some criteria that help in distinguishing them. They conclude by presenting some historical considerations regarding the diachronic relationship between classifiers and nouns in Western Amazonian languages that are relevant for the general theory of classifiers.

Chiara Naccarato and **Shanshan Huang**'s chapter, entitled *Binominals denoting instruments: A contrastive perspective*, focuses on a contrastive analysis of complex nominals referring to instruments pertaining the semantic field of cooking in four typologically distant languages: Italian, Russian, Mandarin Chinese, and Japanese. Adopting an onomasiological approach to word-formation, the authors investigate the morphosyntactic strategies adopted by the four languages to create instrumental nouns, which kinds of instrument are more often expressed through binominal lexemes, and what are the semantic relations between the two constituents. They conclude by observing the existence of a possible correlation between the "onomasiological type" (as proposed by Štekauer 1998 and revised by Pepper 2018b) and the type of instrument denoted. From a formal point of view, Italian and Russian more often use derivational processes (**der**) and adjectival (**adj**) or prepositional (**prp**) constructions, while Chinese and Japanese only employ noun-noun compounding (**cmp**).

Steve Pepper's second contribution, *Hatcher-Bourque: Towards a reusable classification of semantic relations*, introduces the second part of this volume, in which the focus is on the unstated (or underspecified) semantic relation that pertains between the two nominal constituents of a binominal. Pepper proposes a 'reusable' classification based on a synthesis of Anna Granville Hatcher's high-level, four-way classification of 'non-appositional' relations (which he extends to also cover appositional relations) and Yves Bourque's low-level classification of 25 relations (extended to 29 by Pepper). The resulting, two-level 'Hatcher-Bourque' classification is proposed as the basis for further collaborative work in the domain of semantic relations, together with a freely available Excel-based tool for computer-assisted annotation of binominals (Pepper 2021).

László Károly's chapter discusses *Binominal strategies and semantic correlations in Turkic languages* based on data from five different branches of the Turkic family: Turkish, Kazakh, Uigur, Khakas and Yakut. The paper focuses on the question how the derivational strategy (**der**) and compounding (understood broadly as including the **con**, **jxt**, **adj** and **gen** strategies) are related to one another in terms of their semantic capacity, interchangeability and competitiveness. His thorough analysis of 201 semantic concepts in the five languages demonstrates

that compounding is an active and frequently used word-formation strategy in Turkic that may significantly surpass derivation in terms of productivity in certain domains of the lexicon. Some minimal examples show that derivation and compounding are interchangeable, and thus that they are in competition. Károly's application for the first time in Turkic studies of Štekauer's (1998) onomasiological theory of word-formation is found to be an adequate framework for the systematic comparison of derivation and compounding.

The chapter *A classification of compounds in Karachay-Balkar* by **Aslı Gürer** provides an analysis of compounds in an understudied Turkic language belonging to the Kipchak group, Karachay-Balkar. More specifically, the author compares noun-noun compounds (**jxt**) with the *izafet* type, in which a marker occurs on the head of the construction (**con**). The latter is a common strategy in Turkish. The analysis shows that the distribution of the "linking element" (i.e. the *izafet* marker) is not optional but rather signals an argument relation between the head and the dependent, that is, its distribution overlaps with that of "transitive" (i.e. relational) nouns. The analysis furthermore indicates that the **con** strategy tends to be found in endocentric subordinate compounds.

Polina Pleshak's paper on *Binominal lexemes in Moksha and Hill Mari* describes the syntax and semantics of binominal lexemes in two Finno-Ugric languages spoken in the Volga Region: Moksha and Hill Mari. She shows that Moksha and Hill Mari demonstrate competition between two types of nominal modification construction, juxtaposed structures (**jxt**) and genitival constructions (**gen**) that do not express core possessive relations. In addition, she shows that the Finno-Ugric genitive has noncanonical attributive functions in certain contexts, and shares morphosyntactic properties with attributivizers. As regards compounds, she notes that binominals denoting similar relations are present in the dictionary as (fused) one-word and as two-word compounds, suggesting that this difference is not significant. Rather, the difference between more lexicalized one-word and more compositional two-word compounds is more relevant as it affects morphophonology and syntax. Whereas one-word compounds (**cmp**) are strict lexical units, compounds consisting of two or more words (**jxt**) can have more complex syntactic properties, which have to be taken into consideration in the classification.

Maria Rosenberg's chapter *Binominals and potential competitors in language development: Evidence from Swedish* deals with a specific type of binominal lexeme, noun-noun compounds (**cmp**). Employing Swedish production data from five children between 1 and 3 years old and taking an onomasiological approach, the author investigates some potentially competing constructions used to express semantic relations that are usually encoded by binominal lexemes in cross-linguistic perspective. Rosenberg concludes that competing structures have

a narrower semantic scope than noun-noun compounds in Swedish, thus making them the strategy most used by children to express the relevant semantic relations. This can be cognitively explained given that juxtaposition is the preferred option because of its "structural accessibility", while some other structurally more complex strategies (e.g. prepositional structures) are acquired later.

References

Arkadiev, Peter & Marian Klamer. 2019. Morphological theory, language description and typology. In Jenny Audring & Francesca Masini (eds.), *The Oxford handbook of morphological theory*. Oxford: Oxford University Press.

Baldwin, Timothy & Su Nam Kim. 2010. Multiword expressions. In Nitin Indurkhya & Fred J. Damerau (eds.), *Handbook of natural language processing*, 267–292. Boca Raton, FL: CRC Press.

Benveniste, Émile. 1966. Différentes formes de la composition nominale en français. *Bulletin de la Société de Linguistique de Paris* 61(1). 82–95.

Boas, Franz (ed.). 1911. *Handbook of American Indian languages*. Washington, D.C.: Bureau of American Ethnology.

Booij, Geert. 2009. Phrasal names: A constructionist analysis. *Word Structure* 2(2). 219–240. https://doi.org/10.3366/E1750124509000427.

Booij, Geert. 2010. *Construction Morphology*. Oxford: Oxford University Press.

Booij, Geert E., Christian Lehmann & Joachim Mugdan (eds.). 2000. *Morphologie: ein internationales Handbuch zur Flexion und Wortbildung*. Vol. 1. Berlin: Mouton de Gruyter.

Booij, Geert E., Christian Lehmann & Joachim Mugdan (eds.). 2004. *Morphologie: ein internationales Handbuch zur Flexion und Wortbildung*. Vol. 2. Berlin: Mouton de Gruyter.

Brown, Cecil H. 2001. Lexical typology from an anthropological point of view. In Martin Haspelmath, Ekkehard König, Wolfgang Oesterreicher & Wolfgang Raible (eds.), *Language typology and language universals: An international handbook*, 1178–1190. Berlin: Mouton de Gruyter.

Cristofaro, Sonia. 2009. Grammatical categories and relations: Universality vs. language-specificity and construction-specificity. *Language and Linguistics Compass* 3(1). 441–479.

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

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

Dryer, Matthew S. 1997. Are grammatical relations universal? In Joan L. Bybee, John Haiman & Sandra A. Thompson (eds.), *Essays on language function and language type. Dedicated to T. Givón*, 115–143. Amsterdam: John Benjamins.

Dryer, Matthew S. & Martin Haspelmath (eds.). 2013. *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. http://wals.info/ (22 February, 2018).

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.

Goldberg, Adele E. 1995. *Constructions: A construction grammar approach to argument structure*. Chicago: The University of Chicago Press.

Goldberg, Adele E. 2006. *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.

Goldberg, Adele E. 2013. Constructionist approaches. In Thomas Hoffmann & Graeme Trousdale (eds.), *The Oxford handbook of Construction Grammar*. Oxford: Oxford University Press.

Goldberg, Adele E. 2019. *Explain me this*. Princeton: Princeton University Press.

Haspelmath, Martin. 2007. Pre-established categories don't exist: Consequences for language description and typology. *Linguistic Typology* 11(1). 119–132. https://doi.org/10.1515/LINGTY.2007.011.

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.

Hilpert, Martin. 2014. *Construction grammar and its application to English*. Edinburgh: Edinburgh University Press.

Hilpert, Martin & Jan-Ola Östman (eds.). 2014. Reflections on constructions across grammars. *Constructions and frames* 6(2). Special Issue.

Hoffmann, Thomas & Graeme Trousdale (eds.). 2013. *The Oxford handbook of Construction Grammar*. Oxford: Oxford University Press.

Hüning, Matthias & Barbara Schlücker. 2015. Multi-word expressions. In Peter O. Müller, Ingeborg Ohnheiser, Susan Olsen & Franz Rainer (eds.), *Word-formation: An international handbook of the languages of Europe. Vol. 1*, 450–467. Berlin: Mouton de Gruyter.

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.

Koch, Peter. 2001. Lexical typology from a cognitive and linguistic point of view. In Martin Haspelmath, Ekkehard König, Wolfgang Oesterreicher & Wolfgang Raible (eds.), *Language typology and language universals: An international handbook*, 1142–1178. Berlin: Mouton de Gruyter.

Koptjevskaja-Tamm, Maria. 2008. Approaching lexical typology. In Martine Vanhove (ed.), *From polysemy to semantic change: Towards a typology of lexical semantic associations*, 3–52. Amsterdam: John Benjamins.

Koptjevskaja-Tamm, Maria, Ekaterina Rakhilina & Martine Vanhove. 2015. The semantics of lexical typology. In Nick Riemer (ed.), *The Routledge handbook of semantics*, 434–454. Abingdon: Routledge.

Koptjevskaja-Tamm, Maria & Ljuba N. Veselinova. 2020. Lexical Typology in morphology. In *Oxford research encyclopedia of linguistics*. Oxford: Oxford University Press. https://doi.org/10.1093/acrefore/9780199384655.013.522.

Levi, Judith N. 1976. A semantic analysis of Hebrew complex nominals. In Peter Cole (ed.), *Studies in modern Hebrew syntax and semantics: The transformational-generative approach*, 9–55. Amsterdam: North-Holland.

Lieber, Rochelle, Sabine Arndt-Lappe, Antonio Fábregas, Christina L. Gagné & Francesca Masini (eds.). 2021. *The Oxford encyclopedia of morphology*. 3 vols. Oxford: Oxford University Press.

Lieber, Rochelle & Pavol Štekauer (eds.). 2009a. *The Oxford handbook of compounding*. Oxford: Oxford University Press.

Lieber, Rochelle & Pavol Štekauer. 2009b. Introduction: Status and definition of compounding. In Rochelle Lieber & Pavol Štekauer (eds.), *The Oxford handbook of compounding*, 3–18. Oxford: Oxford University Press.

Lieber, Rochelle & Pavol Štekauer (eds.). 2014. *The Oxford handbook of derivational morphology*. Oxford: Oxford University Press.

Martinet, André. 1949. La double articulation linguistique. *Travaux du Cercle linguistique de Copenhague* 5. 30–37.

Masini, Francesca. 2009. Phrasal lexemes, compounds and phrases: A constructionist perspective. *Word Structure* 2(2). 254–271.

Masini, Francesca. 2019a. Multi-word expressions and morphology. In *Oxford research encyclopedia of linguistics*. Oxford: Oxford University Press. https://doi.org/10.1093/acrefore/9780199384655.013.611.

Masini, Francesca. 2019b. Competition between morphological words and multiword expressions. In Franz Rainer, Francesco Gardani, Wolfgang U. Dressler & Hans Christian Luschützky (eds.), *Competition in inflection and word-formation,* 281–305. Cham: Springer.

Masini, Francesca & Jenny Audring. 2019. Construction morphology. In Jenny Audring & Francesca Masini (eds.), *The Oxford handbook of morphological theory*, 365–389. Oxford: Oxford University Press.

Michaelis, Susanne Maria, Philippe Maurer, Martin Haspelmath & Magnus Huber (eds.). 2013. *APiCS Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. https://apics-online.info/.

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

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

Pepper, Steve. 2018a. Lexico-constructional patterns in binominal lexemes. Paper presented at SLE Summer School 2018, Tartu. https://www.academia.edu/50523906/.

Pepper, Steve. 2018b. Onomasiological types and the typology of binominal lexemes. Paper presented at the *Typology and Universals in Word-Formation IV* conference, Košice. 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. 2021. *The Bourquifier: An application for applying the Hatcher-Bourque classification*. MS Excel. http://folk.uio.no/stevepe/Bourquifier.xlsx.

Pepper, Steve. This volume, a. Defining and typologizing binominal lexemes. In Steve Pepper, Francesca Masini & Simone Mattiola (eds.), *Binominal lexemes in cross-linguistic perspective*. Berlin: Mouton de Gruyter.

Pepper, Steve. This volume, b. Hatcher-Bourque: Towards a reusable classification of semantic relations. In Steve Pepper, Francesca Masini & Simone Mattiola (eds.), *Binominal lexemes in cross-linguistic perspective*. Berlin: Mouton de Gruyter.

Ramat, Paolo. 1990. Definizione di "parola" e sua tipologia. In Monica Berretta, Piera Molinelli & Ada Valentini (eds.), *Parallela 4: Morfologia, Atti del V incontro Italo – Austriaco della Società di linguistica italiana, Bergamo ottobre 2–4, 1989*, 3–15. Tübingen: Narr.

Rubino, Carl. 2013. Reduplication. In Matthew S. Dryer & Martin Haspelmath (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. http://wals.info/chapter/27 (22 February, 2018).

Štekauer, Pavol. 1998. *An onomasiological theory of English word-formation*. Amsterdam: John Benjamins.

Štekauer, Pavol, Salvador Valera & Lívia Körtvélyessy. 2012. *Word-formation in the world's languages: A typological survey*. Cambridge: Cambridge University Press.

Talmy, Leonard. 2007. Lexical typologies. In Timothy Shopen (ed.), *Language typology and syntactic description. Vol. 3: Grammatical categories and the lexicon*, 66–168. 2nd ed. Cambridge: Cambridge University Press.

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.

Vanhove, Martine (ed.). 2008. *From polysemy to semantic change: Towards a typology of lexical semantic associations*. Amsterdam: John Benjamins.

Wojtylak, Katarzyna I. 2017. *A Grammar of Murui (Bue): A Witotoan language of Northwest Amazonia*. Cairns: James Cook University PhD dissertation.