Compounds and multi-word expressions in Finnish

1 Introduction

Most of the processes to expand the vocabulary of a language are based on a recycling principle: Instead of creating not yet occupied arbitrary sound sequences for new concepts, existing lexemes or morphemes are reused as material for new words. This can happen by borrowing a word from some other language or by altering the meaning and thus shifting the extension of an existing word. Yet, these means are fairly unsystematic. Instead, a system of word-formation offers productive models for expanding the lexicon in an economic way, and it is actually the most common way it happens.1

Word-formation types such as (1a–f) are usually regarded as a domain of morphology:

(1a) Composition (combining lexemes into a new lexical item):
    *kesä* ‘summer’ + *yö* ‘night’ > *kesäyö* ‘summer night’

(1b) Derivation (adding an affix):
    *kesä* ‘summer’ + *-inen* (adjectival suffix) > *kesäinen* ‘summery’

(1c) Backformation (removing an actual or supposed affix):
    *tarrata* ‘grab, stick’ > *tarra* ‘sticker’

(1d) Conversion, also called zero derivation (functional shift of a word or a stem2 without adding morphological material):
    *minä* ‘I’ (Pron) > *minä* ‘ego’ (N); *painia* ‘wrestle’: *paini*- (verb stem) > *paini* (N) ‘wrestling’

---

1 Foreign influence can manifest itself in word formation, too, as calques of singular formations or by taking over a formation model from another language. Many Finnish compounds are loan translations from (or via) Swedish or German. Nowadays loan translations come increasingly from English, cf. *jakamis-talous* < *sharing economy*, *palvelu-muotoilu* < *service design*. In terminology, neoclassical compounds (with elements from Greek or Latin) as internationalisms play an important role.

2 The word stem is the form to which affixes can be attached. As for word stems in Finnish, cf. ISK (2004: 86–89).
Blending (merging parts of existing lexemes combining their semantic features): 

\( \text{kamraati} \) ‘comrade’ + \( \text{toveri} \) ‘companion, friend’ > \( \text{kaveri} \) ‘friend, mate’

Clipping (shortening a lexeme without changing the meaning):

\( \text{akkumulaattori} \) > \( \text{akku} \) ‘accumulator’; \( \text{informaatioteknologia} \) > IT ‘information technology’; \( \text{sosiaaliturva} \) > \( \text{sotu} \) ‘social security’

However, also syntactic (phrasal) sequences can be lexicalized as nominations of specific concepts. Such multi-word expressions (MWEs) can be included in a discussion of word formation in a broad sense. MWEs are fixed word-groups with lexical, syntactic, semantic, pragmatic and/or statistical idiosyncrasies (Sag et al. 2002; Baldwin/Kim 2010; Hüning/Schlücker 2015). The term “multi-word expression” is established above all in computational linguistics; traditionally MWEs are called “phrasemes” or “idioms”. In this chapter, the term “idiom” is used for semantically idiosyncratic MWEs only, i.e. for cases where the meaning of an MWE cannot be concluded from the meanings of its components. MWEs can be fully idiomatic (2a–b), semi-idiomatic (2c) or non-idiomatic but statistically significant (institutionalized) (2d–e):

\[ \text{(2a)} \quad \text{mennä mönkäään} \ (\text{lit. go unique component}) \ ‘\text{go wrong}' \]

\[ \text{(2b)} \quad \text{musta hevonen} \ (\text{lit. black horse}) \ ‘\text{dark horse}' \ (\text{a little known candidate or competitor who unexpectedly wins or succeeds}) \]

\[ \text{(2c)} \quad \text{valkoinen valhe} \ ‘\text{white lie}' \ (\text{a harmless lie}) \]

\[ \text{(2d)} \quad \text{rauhanomainen rinnakkaiselo} \ ‘\text{peaceful coexistence}' \ (\text{theory of the Soviet Union about relations between socialist and capitalist states during the Cold War}) \]

\[ \text{(2e)} \quad \text{neoliittinen kausi} \ (\text{altering with the compound } \text{neoliitti+kausi}) \ ‘\text{Neolithic Period}' \]

---

3 An overview of phraseology with examples from several European languages, e.g. German, English, French, Swedish and Finnish, is given by Korhonen (2018).

4 In Finnish, non-idiomatic MWEs have been studied primarily in terminology with a focus on nominal terms. It can be assumed that ongoing studies in computational linguistics will shed more light on the proportion of non-idiomatic MWEs in standard language, too.

5 In the examples, the compound constituent boundaries are marked with “+”, if needed. Occasionally Finnish case form abbreviations are used as subscripts: \( \text{ALL} = \text{allative}, \text{ELAT} = \text{elative}, \text{GEN} = \text{genitive}, \text{ILL} = \text{illative}, \text{INESS} = \text{inessive}, \text{PART} = \text{partitive} \).
The boundaries between different formation types are not always clear-cut: Compound nouns often compete with MWEs, for example as constructional synonyms in terminology, cf. (2e) above. Some Finnish compounds have internal inflectional elements, which is a syntactic feature (cf. Section 2.1). Moreover, there are hybrid formations, like the so-called “derived compounds” (Section 2.3.1.1, group 2). And finally, scholars have divergent views of certain structures, such as Finnish particle verbs that have been classified either as compounds, prefix derivations or MWEs (Section 3).

Compounds and MWEs share some characteristics: Both are complex lexical units and thus secondary signs for a specific concept, their constituents are words, and they can bear an idiomatic (figurative or opaque) or non-idiomatic (transparent) meaning. One instance of opaqueness is presented by unique components (isolates, cranberry morphemes), compare the MWE in (2a) with the cranberry-compound *puna+tulkku* (lit. red+UNIQUE COMPONENT) ‘bullfinch’ (cf. Nenonen 2002: 13, 15, 21f., 37–40; Stein 2012: 227f.). Both compounds and MWEs can express determinative, appositive and coordinative relations. The compound constituents occur in a fixed order; regarding MWEs this applies mainly to nominal, adjectival and adverbial expressions, whereas verbal MWEs are more flexible. In Finnish, the great majority of compounds are nouns (N), while among idiomatic MWEs verb idioms (V) are the predominant class.

In this chapter, the focus is on the characteristics of compounds, with remarks on differences and overlap in the structure and syntactic distribution of compounds and (fixed or free) phrasal units. Section 2 gives an overview of compounding in Finnish, mostly using examples of nouns and adjectives;\(^6\) In Section 2.1 characteristics of prototypical compounds and their absence, making a compound less prototypical and bringing it nearer to an MWE, are discussed. Section 2.2 deals with the complexity of compounds, and in Section 2.3 the main semantic-hierarchical and morphosyntactic types of compounds are presented. Section 3 focuses on a word class that has been regarded as rather peripheral from the perspective of compounding in Finnish, namely complex verbs. They are interesting for two reasons: They are on the increase in modern Finnish, and they lie at the intersection of compounds (3.1), prefix derivatives (3.2) and MWEs (3.3). In the closing remarks (Section 4) observations on the blurred border between Finnish compounds and MWEs are gathered and suggestions for future research are presented.

\(^6\) Due to lack of space, a thorough description of all word classes is impossible.
2 Compounding in Finnish

2.1 Prototypical compounds

Finnish has an extensive system of word-formation: Both derivation and compounding are highly productive. In particular the diversity and productivity of suffix derivation is often regarded as a special characteristic of Finnish, but, actually, the majority of new words in modern Finnish are compounds (cf. Tyysteri 2015: 13, 223). Verbs, however, show a different profile: There is a rich and productive suffixation system, whereas compounding plays a marginal role. Yet, in the last decades the number of compound verbs has increased.

A compound is a combination of two or more lexemes constituting a new, complex word with a new lexical-conceptual meaning that is generally more specific than the additive meaning of its parts, e.g. märkä+puku ‘wetsuit’ (water sports garment) vs. märkä puku ‘wet suit’. The constituents can be simplex lexemes, derivatives or even compounds, i.e. compounding is potentially recursive. In contrast to derivatives, vowel harmony (cf. Karlsson 2015: 16ff.) does not extent over the constituent boundary (Koivisto 2013: 170), i.e. the integration grade of compounds is lower, compare the suffix derivatives with vowel harmony in (3a) with the compounds in (3b):

(3a) Verb stem + suffix -jA → juoja ‘drinker’ vs. syöjä ‘eater’
(3b) yö+juna (*yö+jynä) ‘night train’, varpus+pöllö (*varpus+pollo)
    (lit. sparrow+owl) ‘pygmy owl’

The main characteristics of prototypical compounds in Finnish are: 1) The constituents occur also as autonomous lexemes, 2) the boundary between the constituents corresponds to a syntactic boundary, 3) the compound has only one main stress that – just as in simplex words – is on the first syllable, 4) a formally identical phrasal unit is not possible, 5) semantically, the compound has become estranged from the meanings of its constituents and lexicalized into a nomination of a concept of its own, 6) morphologically, the compound is internally invariable. Among new compounds in present-day Finnish the proportion of prototypical compounds is increasing, whereas non-prototypical features accumulate on one and the same words. However, counter to the trend towards prototypicality, formations with a non-autonomous pre-element (cf. criterion 1) are on the advance. Further, deriving new verbs and adjectives from already existing compounds, which leads to secondary “derived compounds” where the constituent boundary deviates from the logical syntactic-semantic structure (cf. criterion 2
and Section 2.3.1.1, group 2), has become more common than earlier (Tyysteri 2015).

As a general rule, Finnish compounds are written without space between the constituents, cf. (4) below. Hyphenation is obligatory in case of hiatus (5a) and to indicate the constituent boundary after a special sign (letter, number, acronym etc.) (5b). A compound differs also prosodically from a phrase: The main stress is on the first compound constituent (cf. criterion 3), while in a corresponding phrase both words have a stress of their own (Pääkkönen 1989: 371; Vesikansa 1989: 213; ISK 2004: 388), cf. (6). Yet, stress is not a reliable criterion: Adverbial and conjunctional units (7) bear only one main stress on the first part and show a strong tendency towards univerbation. Until the 1960s they could be written together or apart, today the orthographical norm requires separation and thus an MWE status for them, which is in contradiction with the stress pattern (Niinimäki 1992).

(4)  *metsäyhö* (lit. forest+company) ‘forestry company’

(5a)  *öljy-yhtö* ‘oil company’

(5b)  *A-vitamiini* ‘vitamin A’

(6)  *músta+rastas* (lit. black+thrush) ‘blackbird’ vs. *músta rástas* ‘black thrush’

(7)  *sítä vastoin* (lit. itPART against) ‘by contrast’  
     *niin ollin* (lit. so being) ‘thus, hence’  
     *niin kuin* (lit. so as) ‘as, as if’

Generally, adjectival compounds can have descriptive, graduating or evaluative modifiers in the genitive; semantically relative adjectives like *kokoinen* ‘of the size of’, *näköinen* ‘looking like’ etc. even require a complement in the genitive. Here compounds and phrasal units of identical parts are often interchangeable (cf. criteria 4 and 5), as illustrated in (8). Generally, conventionalized (especially idiomatized) combinations and those with short components undergo univerbation, but a grey zone remains, cf. (9a) and (10a) vs. (9b) and (10b).

(8)  *sydämen+muotoinen* ~ *sydämen muotoinen* ‘heart-shaped’  
     *vaalean-vihreä* ~ *vaalean vihreä* ‘light green’  
     *hassun+kirjava* ~ *hassun kirjava* ‘kooky colorful’

(9a)  *ruohon+vihreä* ‘grass-green’ (conventional)

(9b)  *pinaatin+vihreä* ~ *pinaatin vihreä* ‘spinach green’ (occasional)
In contrast, the first constituent of similitative adjectives that expresses an entity for which the property denoted by the head is typical, is – regardless of its nominative or genitive form – always unified with the head, cf. (11a–b). Here, alternation with multi-word similes depends on syntactic distribution: Similative compounds can be replaced by phrasal similes in predicative (12a) and adverbial function but not in attributive function (12b). They cannot always be exchanged, though: While similitative compounds are mostly lexicalized stereotypes and the first constituent cannot have its own qualifiers, the expression potential of phrasal similes is broader: They are based on a productive phraseosyntactic pattern that is filled with conventionalized (lexicalized) or occasional word combinations, and the component that denotes point of comparison can have supplementary expansions, cf. (13a–b). Thus similitative adjectives and phrasal similes – both typically used for intensification – are partly in a competitive, partly in a complementary relation to each other (ISK 2004: 411; Heinonen 2010).

(11a) jää+kylmä ‘ice-cold’
(11b) langan+laiha (lit. thread +thin) ‘sceletous’
(12a) Koir a oli salaman+nopea (lit. lightning +quick) ~ Koir a oli nopea kuin salama (lit. quick as lightning). ‘The dog was as quick as a lightning.’
(12b) salamannessa koira ~ *nopea kuin salama koira
(13a) hidas kuin etana ‘slow as a snail’ (conventional idiom) ~ etana+hidas
(lit. snail +slow)
(13b) hidas kuin halvaantumut etana ‘slow as a paralyzed snail’ (occasional expansion) ~ *halvaantuneen+etana+hidas (lit. paralyzed snail +slow)

Morphological integrity of prototypical compounds means that they are internally invariable (cf. criterion 6); the morphological head bears the inflectional elements. However, in some compounds the adjectival first constituent can (14a) or must (14b) agree in case and number with the nominal head, which indicates their phrasal origin (cf. Niemi 2009: 239 f.):

(14a) iso+sisko ‘big sister’ – allative: isolle+siskoalle ~ iso+siskoalle
(14b) nuori+pari (lit. young+couple) ‘newlyweds’ – allative: nuorelle+parille
Internal congruence is a recessive property. Of the 587 A+N compounds in “Suomen kielen perussanakirja” (Basic dictionary of Finnish, 1990–1994) 84% do not allow congruence, while the remaining 16% are distributed fairly equally among compounds with obligatory vs. optional congruence. In neologisms and occasional compounds non-congruent first constituents are almost exclusive (ISK 2004: 392, 406; Tyysteri 2015: 141–148). A compound without internal inflection underlines the term character: *oma+lääkäri* is a personal doctor nominated for a certain patient by public health care (15a). In contrast, the corresponding (congruent) attributive NP refers to a non-administrative private choice made by the patient (15b):

15a) *oma+lääkäri* (lit. own+doctor) – allative: *oma+lääkärille*
15b) *oma lääkäri* ‘own doctor’ – allative: *omalle lääkärille*

However, a special class of compounds with internal inflection remains: compound numerals (16a). To avoid overlong compounds, numerals with hundreds, thousands etc. are “cut” into groups (ISK 2004: 756f.) so that they combine features of MWEs and compounds (16b).

16a) *kolmelle+kymmenelle+neljälle* ‘34ALL’
16b) *kahdelle+kymmenelle+tuhannelle seitsemälle+sadalle kolmelle+kymmenelle+neljälle* ‘20734ALL’

2.2 Complexity of compounds

The majority of Finnish compounds consist of nominal compounds. The most common type is a combination of two base (i.e. non-derived) nouns (N+N), the largest group being determinative compounds (cf. Section 2.3.1.1) with the first constituent in the (endingless) nominative case (Karlsson 2015: 282; Pitkänen-Heikkilä 2016: 3213). The typical base word structure in Finnish is bisyllabic, so even compounds with two base words have mostly at least four syllables (17a), and since derivatives and compounds can function as compound constituents as well, Finnish compounds tend to be long (Karlsson 2004: 1329), cf. (17b). In principle, there is no upper limit on the complexity, but increasing complexity diminishes intelligibility. As a consequence of recursiveness, compounds with four or five components are not rare in languages for special purposes (e.g. administra-

---

7 There are less than 100 monosyllabic word roots in Finnish, whereas English has at least 7,000 (Karlsson 2004: 1329).
tion, medicine etc.), yet, (mostly occasional) polymorphemic compounds appear also in everyday language (17c). In Tyysteri’s corpus\(^8\) two-constituent compounds dominated with 83.6\%, whereas the ratio of three-constituent compounds ran into 15.5\% and that one of four-constituent compounds into 0.9\%; longer formations occurred only sporadically (Tyysteri 2015: 100–104; as for letter number in compounds, cf. ibid.: 104–108).

(17a) vesi+pullo 'water bottle'
(17b) työ+ehto+sopimus+neuvottelut (lit. work+condition+contract+negotiations) 'negotiations for collective bargaining’
(17c) peruna+sose+hiutale+pakkaus (lit. potato+mash+flake+package) ‘package of mashed-potato flakes’

2.3 Main types of compounds

2.3.1 Semantic-hierarchical structure

Like in many other languages, Finnish compounds can be categorized as either determinative (subordinate) or copulative (co-ordinate) compounds.

2.3.1.1 Determinative compounds

In determinative compounds the final constituent is the morphosyntactic and semantic head: It bears the inflectional elements and expresses a general concept that is modified by the initial constituent so that the compound denotes a subordinate concept (hyponym) to the head (18a). Such compounds are called endocentric (Olsen 2015: 365f., 370). The modifier is not referential but has a general meaning, which makes the compound semantically different from a corresponding phrase (ISK 2004: 390), cf. (18b). Whether the first constituent is morphologically underspecified (18a) or has a case ending explicating the syntactic relation between head and modifier, cf. (18b), varies from compound to compound.

(18a) kivi+talo ‘stone house’ (a special kind of house: ‘house made of stone’)
(18b) kirkon+kello (lit. church\(^{\text{GEN}}\)+bell) ‘church bell’ vs. (läheisen) kirkon kello ‘bell of the (nearby) church’

---

\(^8\) Tyysteri’s material consists of more than 28.000 new compounds (types) in Finnish print media in the period 2000–2009, collected from Nykysuomen sanastotietokanta (Lexical database of modern Finnish) of Kotimaisten kielten keskus (Institute for the Languages of Finland) (Tyysteri 2015: 79–84).
In Finnish grammar, the following special types are regarded as subclasses of determinative compounds:

1) In **synthetic compounds** the first constituent is comparable with the subject (19a), object (19b) or some other argument (19c) of the verb from which the head is derived (cf. ISK 2004: 400f.; Olsen 2015: 370f.). The first constituent typically has a case ending, which is a syntactic feature transmitted by the verb. Nominizations with -minen are not univerbated with the verb arguments (20a), while deverbal nouns with other suffixes form compounds as well as phrasal NPs (20b).

(19a) *auringon+nousu* (lit. sun_{GEN}+rise) ‘sunrise’
(19b) *kirjan+sitoja* (lit. book_{GEN}+binder) ‘bookbinder’
(19c) *kirkossa+kävijä* (lit. church_{INESS}+goer) ‘churchgoer’

(20a) *pyykin peseminen* (lit. laundry_{GEN}, washing) ‘washing laundry’ vs. *pyykin+peseminen*
(20b) *pyykin+pesu* (lit. laundry_{GEN}+wash) ~ *pyykin pesu*

2) Words with characteristics of both compounds and derivatives are regarded as secondary “**derived compounds**” (Vesikansa 1989, 213; ISK 2004, 388; Koivisto 2013, 334f.; Pitkänen-Heikkilä 2016, 3211). They can be analyzed as derivatives from complex bases, i.e. compound nouns (21a), adjectives (21b) or phrasal items (21c). Yet, language users tend to reanalyze them, setting the morphological main boundary intuitively as if they were “normal” compounds, even if this does not correspond to the logical syntactic-semantic boundary. In (21a), *perus* ‘base’ does not modify the word *koululainen* ‘pupil’ (e.g. in the sense of ‘typical pupil’). Here, the reanalysis from analogical derivation into analogical compounding (i) gives a kind of short cut to build compounds directly (ii). By generalization the right half of the equation in (ii) becomes model character also in cases where one member is missing in the left half, cf. (21b–c) where *mukaistaa* or *pukuinen* do not occur as autonomous words.

(21a) *perus+koululainen* ‘comprehensive school pupil’ < *perus+koulu*
    (lit. base+school) ‘comprehensive school’
    (i) *koulu* (simplex) : > *koulu + -lainen* (suffixation)
        *perus+koulu* (compound) > (*perus+koulu*) + -lainen (suffixation)
        > *perus+koululainen* (reanalysis into a compound)
    (ii) *koulu* : *koululainen = perus+koulu : perus+koululainen*
Possessive compounds (bahuvrīhi) such as (22a–b) show a semantic modification similar to determinative compounds, but, due to a metonymic shift, instead of expressing a subcategory of the concept expressed by the final constituent they “identify the intended referent as the possessor of the particularly salient property” they express; i.e. they are exocentric (Olsen 2015: 367; cf. Vesikansa 1989: 250–254; ISK 2004: 409).

(22a) kalju+pää ‘boldhead’
(22b) puna+rinta (lit. red+chest) ‘robin’

Schellbach-Kopra (1964) assumed that bahuvrīhis are decreasing in modern Finnish, but Heinonen (2001) and Malmivaara (2004) demonstrate their productivity: They are used creatively for example in journalistic texts and colloquial speech.

Confix compounds. It is disputable if Finnish has prefixes at all. That is why formations with “prefix-like elements” are subsumed under compounding and not treated as a subclass of affixation or as a word-formation type of its own (cf. Pitkänen-Heikkilä 2016: 3212). The indigenous negation pre-elements epä- ‘un-’ and ei- ‘non-’, cf. (23a), are often called prefixes, but as they consist of a lexical stem (cf. derivatives like evätä ‘refuse, deny’, epäillä ‘doubt, mistrust’; eittämätön ‘undeniable’), the result is very compound-like (ISK 2004: 192). There are many further indigenous “prefix-like elements” that do not occur as autonomous lexemes but have a more or less lexical meaning (ISK 2004: 192f., 393, 402, 415), cf. (23b). Consequently, they can be classified as confixes (cf. Fleischer/Barz 2012: 63f., 107f., 172ff.). As for verbs with confixes, cf. Section 3.1.

(23a) epä+onni ‘bad luck’
epä+suomalainen ‘un-Finnish’
ei-eurooppalainen ‘non-European’

As for the theoretical status of prefixation in the history of linguistics, cf. Olsen (2015: 364f.).
Compo unds and multi-word expressions in Finnish

Similarly to epä-, ei-, foreign negation prefixes, such as dis-, in-, are treated as compound components in Finnish, as well as other foreign prefixes and confixes, e.g. ex-/eks-, pre-, hyper-, mikro-, poly-, neo-, audio-, anglo-, bio-, geo-, psyko- which occur in neoclassical compounds (see Olsen 2015: 374f.), cf. (24a). Some foreign pre-elements can also be combined with indigenous heads (24b) (cf. Sajavaara 1989: 76ff.10; ISK 2004: 192, 394, 402; Pitkänen-Heikkilä 2016: 3214).

(24a) dis+harmonia ‘disharmony’
neo+nataalinen (med.) ‘neonatal’
(24b) anti+sankari ‘anti-hero’
ex-vaimo ‘ex-wife’

5) Appositive compounds describe one particular referent from different perspectives. In contrast to additive (copulative) compounds (cf. Section 2.3.1.2), the constituents do not belong to the same conceptual category, cf. (25a). Even if the constituents are in an appositive relation to each other, a determinative interpretation is possible (ISK 2004, 407f.). In (25b) it is actually the second constituent that modifies semantically the first one, thus having the same function as a postponed apposition (Vesikansa 1989: 223). Further appositive compounds include subsumptive (explicative) compounds (25c) where the second constituent expresses the hyperonym of the first constituent (ISK 2004: 408).

(25a) prinssi+puoliso ‘prince consort’
(25b) puu+vanhus (lit. tree+oldster) ‘old tree’
(25c) veli+mies (lit. brother+man) ‘brother’
perjantai+päivä (lit. Friday+day) ‘(the weekday) Friday’

6) Iterative compounds repeating the same lexeme are productive primarily in informal, playful style of young people; in standard language they are a marginal class. Their main function is emphasis. In N+N reduplications the first constituent expresses the real, prototypical or ideal character of the concept denoted by

---

10 Sajavaara (1989: 79f.) also gives an overview of bound second constituents of neoclassical compounds in Finnish.
the head and implies a contrast (26). As for adjectives, cf. (27), the first constituent is mostly in the genitive \( A_{\text{GEN}} + A \) and functions as an intensifier; the components can be combined as a compound or a phrase without an essential difference in meaning (ISK 2004: 410; Tyysteri 2015: 66f.), similar to (8) above.

(26) \( \text{ruoka}+\text{ruoka} \) (lit. food+food) ‘real food’ (in contrast to fast food or unhealthy food)
\( \text{kirja}+\text{kirja} \) (lit. book+book) ‘printed book’ (in contrast to e-book).

(27) \( \text{pienen}+\text{pieni} \) (lit. small\( \text{GEN} \)+small) \( \sim \) \( \text{pienen pieni} \) ‘tiny, minuscule, itsy-bitsy’

2.3.1.2 Copulative compounds

Copulative compounds consist of two or more parallel (coordinate) parts belonging to the same word class and the same conceptual category; the rightmost constituent is the morphological head.

Historically, co-compound (\textit{dvandva}) is the oldest compound type in the Finno-Ugric languages (Vesikansa 1989: 214; Pitkänen-Heikkilä 2016: 3213). “Co-compounds denote a hyperonym of their constituents, or a superordinate concept” (Olsen 2015: 368); hence, they are exocentric. In Finnish, only \( \text{maa}+\text{ilma} \) (lit. earth+air) ‘world’ has remained until our days.

Additive compounds make up a productive subclass of copulative compounds. Their constituents represent the same conceptual category and stand semantically in an additive relation, similar to members in a syntactic coordination (ISK 2004: 416f.; Pitkänen-Heikkilä 2016: 3213). In Finnish, appositive compounds are dissociated from additive ones orthographically: The former are written as one word, cf. (25a–c) above; the latter are generally written with a hyphen, cf. (28).

(28) \( \text{laulaja-näyttelijä} \) ‘actor-singer’
\( \text{jääkaappi-pakastin} \) ‘fridge-freezer’
\( \text{musta-puna-keltainen} \) ‘black-red-yellow’

2.3.2 Morphosyntactic classification

The primary morphosyntactic classification criterion of Finnish compounds is the word class of the head which determines the word class of the compound. There are no head-based categorical restrictions for the non-head constituent; it
can be a stem, case form or a specific combining form.\footnotemark[11] The first component is usually classified on grounds of its word class (if identifiable) and/or its form (nominative, genitive, other case form, combining form, indeclinable element or element with deficient paradigm). Subclasses that arise from the cross classification of the morphosyntactic types of both constituents are described semantically in detail in the research literature, but no hard and fast rules can be given.

It is a controversial question to what extent the meaning of a compound is influenced by the form of the first constituent. The most frequent first constituent form in Finnish is the nominative which is the base form without any inflectional elements. This base form, as well as the combining forms, leaves the constituent relation underspecified so that several interpretations are possible. Inherently ambiguous compounds can be interpreted semantically and pragmatically, such as world knowledge of prototypical (e.g. local, temporal, causal, instrumental, possessive etc.) relations, common ground and contextual inference (cf. Olsen 2015: 365f., 376 ff., 382; Pitkänen-Heikkilä 2016: 3213). Lexicalized and frequently used compounds can be understood holistically, without analytic compositional processing, but there is psycholinguistic evidence that some form of analysis is co-present (Mäkisalo 2000). Räisänen (1986) points out that lexicalized compounds can be reinterpreted on contextual grounds: In a football report, *maa+pallo* (lit. earth+ball) and *ilma+pallo* (lit. air+ball) with the lexicalized meaning ‘globe’ resp. ‘balloon’ are interpreted in a context-adequate way as occasionalisms describing the motion of the ball either along the ground or through the air.

If the first constituent is in the genitive or some other non-nominative case, the interpretation is more restricted. In such cases the head is usually a deverbal noun and the first component corresponds to an argument of the underlying verb (synthetic compounds, cf. Section 2.3.1.1, group 1). A first constituent in the genitive can indicate a (in a broad sense) subjective-possessive (29a) or objective relation (29b); the latter is more common (Saukkonen 1973: 338; cf. ISK 2004: 400). Locative cases are also current (29c). It is noteworthy, however, that case marking is not obligatory: Similar relations can also be expressed by compounds with morphologically unspecified modifiers (30a–c).

---

\footnotetext[11]{A combining form (casus componens) is a form of the non-head constituent that as such does not occur as an autonomous word form. Besides non-autonomous stem forms, such as *nais- < nainen* ‘woman’ (*nais+ryhmä* ‘women’s group’) or *pien- < pieni* ‘small’ (*pien+teollisuus* ‘small industries’), there are specific combining forms with additional morphological material. For example, verbal first constituents appear mostly in a combining form with -*ma- or -in-* (*istuma+paikka* (lit. sitting+place) ‘seat’, *levin+uuni* ‘baking oven’ (cf. also Tyysteri 2015: 121, 131, 134f.).}
(29a) *ilmaston+muutos* ‘climate change’
*tien+vieri* ‘roadside’

(29b) *puun+hakkaaja* ‘woodcutter’
*ilman+suodatin* ‘air filter’

(29c) *maasta+muutto* (lit. country + moving) ‘emigration’
*tiiiille+pano* (lit. account + put) ‘deposit, payment into an account’

(30a) *terroristi+hyökkäys* ‘terrorist attack’ (subjective)

(30b) *oppilas+valinta* ‘student selection, selection of pupils’ (objective)

(30c) *koti+matka* ‘home journey’ (adverbial: goal)

There are pairs of compounds with a nominative vs. genitive first constituent where the case choice seems more or less arbitrary (31a–b), and others where the difference in meaning is minimal (32a–b). Yet, sometimes there is a clear semantic opposition: (33a) is a specific house, whereas in (33b) the head describes an action and the first constituent in the genitive is the object argument of the underlying verb (cf. Vesikansa 1989, 230–237; ISK 2004, 398–400).

(31a) *kulta+keräys* ‘gold collection, collecting gold’

(31b) *paperin+keräys* (lit. paper + collection) ‘(waste) paper collection’

(32a) *juusto+pala* ‘cheese piece’ (the first component focuses on material)

(32b) *juuston+pala* (lit. cheese + piece) ‘piece of cheese’ (whole to part relation)

(33a) *sauna+rakennus* ‘sauna building’ (a special type of building)

(33b) *saunann+rakennus* (lit. sauna + building) ‘building of a sauna/saunas’

Case marking on the constituent boundary does not contradict the principle of world-knowledge and context-based interpretation, but in giving further information on the relation between the constituents it can exclude alternatives that are possible when the first constituent is unmarked: While the underspecified form *pöytä+tarjoilu* (lit. table + service) can be used in the meaning ‘buffet service, self-service from the table’ (source), the marked form *pöytiin+tarjoilu* (lit. tables + service) precludes this interpretation because the illative ending makes the opposite direction (goal) explicit.
3 Complex verbs in Finnish at the intersection of compounds, prefix derivatives and MWEs

In Finnish, compound verbs are rare.\textsuperscript{12} They belong to the category of determinative compounds;\textsuperscript{13} the first constituent is a noun, adjective, numeral, pronoun, non-autonomous stem or particle (adverb/adposition) (Rahtu 1984: 409–412; ISK 2004: 414f.). Verbs with a particle as first constituent are often replaced by MWEs with the same elements. On the other hand, some first constituents come near to prefixes. Thus, complex verbs can be explored on a scale MWE – compound – prefix derivative.

Modern Finnish has about 250 lexicalized compound verbs with a full paradigm, but the number is increasing (ISK 2004: 414). Additionally, formations with a deficient paradigm (mostly participle forms) are in use, and occasionalisms occur. Compound verbs were banned by Finnish language planning as loan translations for a long time. In the last decades the norm has become more permissive, which can explain the increasing occurrence (cf. Rahtu 1984: 409; Vesikansa 1989: 254–258; Vaittinen 2003: 50; Tyysteri 2015: 40, 154, 220f.).

There are three historical layers of compound verbs in Finnish: The oldest compound verbs, with an adverb as first constituent, are loan translations from the time of the Reformation. In the end of the 19\textsuperscript{th} century a new type, derived from compound nouns, appeared. In the beginning of the 20\textsuperscript{th} century also adjective compounds became derivation bases of verbs (Häkkinen 1987: 10–19; Vaittinen 2003: 47). Also in modern Finnish most of the compound verbs are secondary “derived compounds”, i.e. derivatives or backformations from compound adjectives or nouns, such as (34a–c) (Vesikansa 1989, 256ff.; Tyysteri 2015: 153; cf. also Section 2.3.1.1, group 2). According to ISK (2004: 414f.), most present-day compound verbs are derived from complex adjectives ending on the suffix -inen, cf. (34a). According to Tyysteri (2015: 158, 213), however, the majority of the newest compound verbs go back to compound nouns (34b–c). For the most part new compound verbs have a noun (N) as first component (Tyysteri 2015: 173), which is

---

\textsuperscript{12} According to Saukkonen (1973: 337f.), the proportion of verbs among all compounds in “Nykysuomen sanakirja” (1951–1961) remains at 0,3\%. In Tyysteri’s (2015: 113) corpus their ratio (types) is 1,2\%.

\textsuperscript{13} Copulative compound verbs do not exist in Finnish. Compounds with a verb stem as first constituent are possible, cf. riippu-liitää ‘hang-glide’, but the constituent relation is determinative, not additive. In itku-nauratkaa ‘make cry and laugh’ (Vesikansa 1989: 258) the semantic relation is similar to an additive compound, but the first constituent is a deverbal noun, i.e. the morphological structure is asymmetric.
unsurprising since compound nouns are the most common derivation base, and among these, the structure N+N is predominant.

(34a) \( \text{kaksi} + \text{kielistyä} \) ‘become bilingual’ < \( \text{kaksi} + \text{kielinen} \) (lit. two+lingual) ‘bilingual’

(34b) \( \text{valo} + \text{kuvata} \) ‘photograph’ (V) < \( \text{valo} + \text{kuva} \) (lit. light+picture) ‘photograph’ (N)

(34c) \( \text{koe} + \text{lentää} \) (backformation) ‘test fly’ < \( \text{koe} + \text{lento} \) ‘test flight’

Adverbs, particles and non-autonomous elements can combine directly with verbal heads (Vesikansa 1989: 254ff.). Such preverbs are often called “prefix-like elements” because they are in many respects similar to prefixes in other languages. In Finnish, however, prefixation is untypical (Häkkinen 1994: 488; Kolehmainen 2006: 111, 113). This is why word formation with bound “prefix-like elements” is subsumed under compounding in the Finnish grammar tradition, even if the notion of “prefix-likeness” varies (cf. Tyysteri 2015: 127ff.). In the following, the focus is on verbs with such prefix-like elements.

Kolehmainen (2006) makes a distinction between position fixed bound preverbs, divided into (a) confixes and (b) prefixes, and in contrast to them (c) separable particles in phrasal verbs. Consequently, in each group the word formation status of the verbs is different: in (a) compound (3.1), in (b) prefix derivative (3.2), and in (c) MWE (3.3). In the following, these groups are examined in detail in order to estimate their structural status and productivity.

### 3.1 Confix compounds

Complex words with a prefix-like first constituent that does not occur as an autonomous lexical unit (and thus has an unspecific word class status) are relatively common in modern Finnish. In Tyysteri’s material, including all word classes, they make up 9,3% of all two-constituent compounds; indigenous and foreign pre-elements are roughly equally common. Yet, the word class distribution (e.g. the ratio of verbs) of such formations is not given (cf. Tyysteri 2015, 118ff., 125, 128). The examples in (35a) are lexicalized compounds (cf. Kolehmainen 2006: 115); neologisms and occasionalisms such as (35b) are being used more and more frequently.

(35a) \( \text{edes} + \text{auttaa} \) (lit. forth+help) ‘help, assist, further’

\( \text{jäldi} + \text{kiilottaa} \) (lit. after+polish) ‘polish bright’
Compo unds and multi-word expressions in Finnish

etä + seurustella (lit. distance + go together) ‘have a long-distance relationship’
pika + syödä (lit. quick + eat), ‘eat quickly, eat fast food’
täsmä + leikata (lit. precise + operate) ‘operate/remove precisely’

The “prefix-likeness” of such elements is debatable. The term confix seems more suitable here because in contrast to semantically abstract prefixes, the pre-elements in question still have a more or less clear lexical-conceptual meaning. Historically, they go back to autonomous lexemes; some of them occur today only as bound elements (e.g. epä- ‘un-, non-’; esi- ‘pre-’; etä- ‘long-distance, remote’), some are obsolete or archaic as autonomous words (e.g. lähi- ‘near’; taka- ‘back, rear’; tasa- ‘even, equal’). Others have an autonomous homonym, but the semantic difference is so big that the common origin is not transparent (e.g. edes- ‘further, forth’; etu- ‘fore, forward, front’; jälki- ‘post-, after-’) 14 (Kolehmainen 2006: 113f., 128). Karlsson (1983: 192f.) points out that these elements are semantically similar to nouns and adjectives and calls them lexical “relic morphemes”.15

Moreover, these elements differ from prefixes in their ability to function as derivation bases (cf. esi- ‘pre-’ in the derivate esittää ‘present, put forth, perform’ vs. esi + katsella ‘preview’; more examples in Kolehmainen 2006, 119). They are somewhere between prototypical compound constituents and affixes (ibid.: 118–124). Confix verbs meet the prototypicality criterion 4) (cf. Section 2.1) according to which a form-identical phrase is not possible (*esi katsella, *katsella esi), but since the first constituent is not an autonomous lexeme (criterion 1), they count as non-prototypical compounds.

In spite of the fact that non-autonomous elements can in principle be combined regularly with verbal heads, many of the complex verbs in this group are actually secondary compounds, i.e. derivatives (36a) or backformations (36b) from already existing compounds (see above).16 Many confix verbs have an incomplete paradigm: They are preferably used in infinite forms, especially as adjec-

---

14 In affirmative expressions the autonomous word edes means ‘at least’ in modern Finnish, with negation it has the meaning ‘[not] even’. The noun etu means ‘advantage, benefit’ and the noun jälki ‘track, trace’. In spite of the common etymology, native speakers hardly associate these words with the corresponding pre-elements (Kolehmainen 2006: 114, 126).
15 ISK (2004: 192, 393, 414f.) and Rahtu (1984: 409) characterize them as “prefix-like nominal stems”.
16 In Tyysteri’s random sample of 300 two-constituent-compounds (100 nouns, 100 adjectives and 100 verbs), 75% of the compound verbs (including all kinds of first constituents) were formed by derivation or backformation and only 25% by regular compounding. The ratio of regular compounding is much lower than in previous studies (Tyysteri 2015: 154f., 158).
tive-like participles, which is a transitional phase on the way towards a full paradigm via analogy and generalization. Analogy plays a role in producing new verbs as well: When verbs with a given initial element, e.g. *ala-* ‘sub-’, become more frequent (e.g. *alaotsikoida* ‘subtitle’, *alaluokitella* ‘subclassify’ etc.), the word structure is reanalyzed such that the main constituent boundary is after the pre-element, and not after the complex nominal base, thus as if the verbs were formed regularly via combining *ala-* directly with the verb. In this way, an originally prenominal confix can develop into a preverbal confix, cf. (i), which leads to a symmetric compounding model (ii) that can be generalized, cf. (iii):

\[
\begin{align*}
(36a) & \quad \text{ala+otsikoida} \ ‘subtitle’ \ (V) < \text{ala+otsikko} \ ‘subtitle’ \ (N) \\
& \quad (i) \quad \text{otsikko} \ ‘title’ \ (N) > \text{otsik-} + \cdot\text{oida} \ (\text{suffixation}) \ ‘title’ \ (V) \\
& \quad \quad \quad \text{ala+otsikko} \ ‘subtitle’ \ (\text{compound} \ N) > (\text{ala+otsik-}) + \cdot\text{oida} \ (\text{suffixation}) \ ‘subtitle’ \ (V) > \text{ala+otsikoida} \ (\text{reanalysis into a compound} \ V) \\
& \quad (ii) \quad \text{otsikko} : \text{otsikoida} = \text{ala+otsikko} : \text{ala+otsikoida} \\
& \quad (iii) \quad N : V = x+N : x+V \\
(36b) & \quad \text{esi+pestä} \ ‘prewash’ \ (V) < \text{esi+pesu} \ ‘prewash’ \ (N)
\end{align*}
\]

In Kolehmainen’s assessment (2006: 116f.), given the limited lexical variation in her research material (76 different verbs with 22 indigenous confixes)\(^{17}\) the structure confix+verb plays a minimal role in modern Finnish, i.e. it is not productive. Yet, according to ISK (2004: 414f.), the number of different verbs with *epä-* ‘un-’, *esi-* ‘pre-’, *jälki-* ‘post-’, *pika-* ‘quick, instant’ is increasing, which means that at least these elements are productive. Among the new compounds from the first decade of the 21\(^{st}\) century many more than the above-mentioned bound preverbs are in frequent use – to an extent that proves the productivity of this formation model (Tyysteri 2015: 130). Confix verbs are, however, often stylistically marked: They occur as terms in languages for special purposes; in everyday language and print media occasionalisms are often used playfully (Vesikansa 1989: 257f.; Kolehmainen 2006: 116; Tyysteri 2015: 88, 113, 213). Nevertheless, it is evident that the number of lexicalized confix verbs in standard language is increasing. The currently most popular indigenous and foreign verb confixes have a high communicative and cultural relevance: They reflect modern life with its hectic pace (*pika-*), green values (*bio-, eko*) and technological innovations (*digi-, nano-, etä-, täsmä-*).

---

17 Kolehmainen collected her research material from dictionaries and authentic texts from the 1990es in SKTP (*Suomen kielen tekstipankki / Language Bank of Finland*).
3.2 Prefix verbs?

The question is whether adpositional and adverbial elements that are used as bound preverbs in Finnish can be regarded as prefixes. Kolehmainen (2006: 130–137) cautiously refers to them as “prefix-like elements” and underlines that they differ in some aspects from prefixes in Germanic languages. Firstly, they are not unstressed: The main word stress in Finnish is generally on the initial syllable, i.e. word stress does not apply as prefix criterion in Finnish. Secondly, the Finnish adpositions are mainly postpositions.18 Thirdly, many of them are secondary adpositions, having developed from inflected forms of relative nouns,19 and have therefore (fossilized) case endings; some of them have a restricted nominal paradigm in several (still existing or historical), mostly locative, especially directional cases. The same holds for adverbs: Many elements occur both as adpositions and as adverbs (ISK 2004: 664 ff.; Tyysteri 2015: 121). Consequently, there are hundreds of different adposition and adverb forms in Finnish, but not all of them function as preverbs.

Kolehmainen’s research material from grammars, previous studies and dictionaries contains 70 of such elements (Kolehmainen 2006: 134–137). “Nyky-suomen sanakirja” (Dictionary of modern Finnish, 1951–1961) mentions 251 complex verbs with these elements, but many of them are marked as archaic, e.g. alas+astua ‘step down’, and almost a half of them occur in univerbated form only as participles, cf. yhteen+laskettu (lit. together+counted) ‘combined’. In both cases separated alternatives are recommended, cf. astua alas, yhteen laskettu (cf. Section 3.3). Thus the number of inseparable verbs in active use is much lower. Most of the elements combine only with one or two verbs (ibid.: 138). About ten elements show a somewhat broader spectrum, e.g. irti- ‘loose, off’, läpi- ‘through, throughout’, sisään- ‘in, inside’, ulos- ‘out, outside’, yli- ‘over’ (ibid.: 137 ff.). All in all, Kolehmainen regards the model prefix+verb as unproductive.

Finnish inseparable verbs of this group are historical relics that go back to old loan translations from Germanic and classical languages resp. to an interference-based formation model (cf. Öhmann 1957: 33ff.; Vaittinen 2003; Toropainen 2017: 72). In Old Literary Finnish (1540–1810) the majority of printed texts were

---

18 In principle, postpositions (and adverbs) can develop into prefixes in SOV-languages where complements precede the verb. SOV is supposed to be the basic word order in Uralic languages; in Finnish, however, the order has changed into SVO. This is one possible explanation for the weak affinity to prefixes. As for typological theories of linearization in connection with prefixes, see the overview in Kolehmainen (2006: 149–156).
19 This is the first step of a gradual grammaticalization called “noun-to-affix-cline”, cf. Lehmann (1985: 304); Hopper/Traugott (2003: 110); as for Finnish Jaakola (1997: 126 ff., 134).
translations of religious texts, following faithfully the formulations in the original (Häkkinen 1994: 11f.). For example Mikael Agricola (about 1510–1557), the “Finnish Luther”, used 810 different compound verbs (including all first element categories) in his texts, which makes up 32,5% of all his compounds on type level (Toropainen 2017: 53, 55, 66, 74). In about 80% of Agricola’s verb compounds the first constituent was an adverb (Häkkinen 1987: 10). In the 17th century such compounds were often replaced with MWEs consisting of a verb and an adverb by Agricola’s successors. In the 19th and 20th centuries compound verbs were combated by purist language planners as un-Finnish or ungrammatical (Häkkinen 1987: 7), resulting in a radical decline of use.

In modern Finnish, most combinations of adverb and verb, such as pois ‘away’ + sulkea ‘close’, are generally recommended to be formed as two separate words, i.e. as MWEs (e.g. by “Kielitoimiston sanakirja” (2006), a dictionary of Standard Finnish), where the adverb is postponed in case of neutral word order, cf. (37a). Yet, in attributive participles the only possible position for the adverb is before the verb. Although such a word order usually promotes univerbation, the norm of writing separately holds for most participles, cf. (37b), even if language users tend to write the parts together. However, when pois precedes an infinitive, the components are written together, in contrast to the reversed order, cf. (37c). The verb irti+sanoa (lit. off+say) ‘discharge, fire; cancel, (fig.) break off’ behaves in some details differently. As for the infinitive, the alternatives are the same (38a), but in passive past participle, the preceding adverb is not separable (38b).

In other words, the rules differ from verb to verb. Some lexicalized verbs cannot be separated at all (39). In some cases separation is combined with semantic difference: In a concrete meaning the adverb is separated (40a), whereas univerbation is preferred in an abstract meaning (40b) (as for orthographical norm, cf. Pääkkönen 1989: 375; Eronen 1996; Tyysteri 2015: 38).

(37a) pois+sulkea, better sulkea pois ‘exclude, rule out’
(37b) pois suljettu vaihtoehto ‘excluded alternative’
(37c) Mitään vaihtoehtoa ei pitäisi pois+sulkea ~ *pois sulkea ~ sulkea pois
‘None of the alternatives should be excluded.’

20 According to Jussila (1988), about 61% of Agricola’s vocabulary has remained in use up to date, but as for compounds, the proportion is only 15,9%; the strongest decline concerns compound verbs.

21 Although infinitive forms preceded by an adverb are normally written together (cf. *pois sulkea, *irti sanoa), the components must be separated, if an enclitic particle, e.g. -kAAn ‘[not] even, anyway, after all’, is appended to the adverb, cf. Ei sitä voi poiskaan sulkea ‘After all, it cannot be excluded’; Ei heitä voi irtikään sanoa ‘Anyway, they cannot be fired’.
In my opinion, these pre-elements are not prefixes. One reason is their obvious unproductivity, i.e. the restricted verb variation per pre-element – for affixes a far wider use is expected. The still existing bound forms are sporadic historical relics, based on calques from foreign languages with systematic prefixation, yet, in Finnish, a generalization never took place. The initial word stress protects the elements from phonological erosion typical of affixes. Above all, the fact that there are parallel phrasal forms, cf. (37a) and (38a), is a proof of the lexical autonomy of the elements in question – in that respect they show a higher autonomy than confixes (cf. Section 3.1). It follows that the univerbated forms are compounds. Here I agree with Tyysteri (2015: 119, 121) who, in contrast to Kolehmainen (2006), does not classify the above-mentioned elements as prefixes or “prefix-like elements” but as “indeclinable elements or elements with incomplete declination (adverbs, adpositions and particles)” in ordinary compounds. The advantage of this analysis is that the coexistence of occurrences with and without separation, i.e. MWEs vs. compounds, can be compared with similar cases in other word classes where both alternatives have (nearly) the same meaning, cf. (8) and (20b) above.

Whether the one-word and the two-word combination represent one and the same verb lexeme or two synonymous lexemes and whether the phrasal alternatives should be regarded as regular (“free”) syntactic constructions or rather as phrasal verbs, i.e. MWEs, is discussed in the next section.

3.3 Phrasal verbs

In the linguistic literature the terms “phrasal verb” and “particle verb” are often used as synonyms. The former implies that the components are separate, while the latter refers to the functional category of the component the verb is connected
with. In English, for example, particle verbs are always phrasal verbs. In Finnish this need not be the case.

In traditional Finnish grammar phrasal verbs are not recognized as an established category, but several scholars refer to fixed sayings or idiomatic figures of speech in the form of MWEs, similar to separable particle verbs in Germanic languages (cf. Häkkinen 1997: 44; Nenonen 2002: 55), cf. (41a). They are semantically and structurally similar to verb idioms consisting of a verb and a non-particle component, for example a unique component (41b) or a nominal component in a locative case (41c) (cf. Nenonen 2002: 55f.; Kolehmainen 2006: 164).

(41a) *panna vastaan* (lit. put against) ‘resist, struggle against’
(41b) *lyödä laimin* (lit. hit/beat UNIQUE COMPONENT) ‘neglect, abdicate’
(41c) *ottaa huomioon* (lit. take account) ‘take into account’

According also to ISK (2004: 447), particle verbs are “idiomatic predicates”. Here “particle” refers to the functional category of the element co-occurring with the verb, regardless of univerbation or separation. In some cases “Kielitoimiston sanakirja” (2006) lemmatizes the univerbated form but refers to the phrasal one. From entries like (42a) it can be inferred that both forms are regarded as representations of the same lexeme; remarks such as ‘mostly’ or ‘better’ (42b) indicate that the MWE is generally the dominant form. Occasionally only the univerbated form is given although separated forms occur commonly, cf. (42c). However, as mentioned above, some verbs are used only in the univerbated form, cf. (39).

(42a) *irti+sanoa* = *sanoa irti*, cf. (38a)
    *laimin+lyödä* = *lyödä laimin*, cf. (41b)
(42b) *ylen+antaa*, mostly *antaa ylen* ‘throw up, vomit’
    *pois+sulkea*, better *sulkea pois*, cf. (37a)
(42c) *ulos+liputtaa* ‘flag out’ vs.
    *Viking Line liputtaa ulos kaksi alusta*. ‘Viking Line is going to flag out two ships.’

Kolehmainen (2006: 170f.) sees the separation (i.e. the MWE structure) and the idiomaticity or metaphoricity of the combination to be key criteria; in her assessment particle verbs are either singular idioms or go back to phraseological patterns. This means that transparent (non-idiomatic) combinations, such as (43), are excluded from the class of particle verbs and regarded as products of free syntax; according to Kolehmainen (ibid.) native speakers do not perceive them as single semantic units.
Yet, it is not always easy to draw the line between idiomatic and free combinations because idiomaticity is a continuum. Kolehmainen (ibid.: 172–183) distinguishes between four grades of idiomaticity and compositionality:

(A) Fully idiomatic combinations that do not permit any component variation are obvious verb idioms, e.g. lyödä laimin, cf. (41b) above, where laimin is a unique (adverb-like) component and the verb lyödä does not bear its regular meaning ‘hit, beat’, cf. (44a) vs. (44b).

(44a) He lyövät laimin lapsiaan. ‘They neglect their children.’ (idiomatic meaning) vs.
(44b) He lyövät lapsiaan. ‘They are beating their children.’ (regular meaning)

Also a combination of verb and autonomous adverb can in principle become fixed as a single idiomatic MWE without component variation, e.g. (45a), where, however, the figurative meaning is compositional to some degree, as far as ampua is understood as a destructive action; the directionality of the adverb underlines telicity (‘once for all’), and in up-down-metaphors ‘down’ means negative things, here (a change into) non-existence. A similar compositionality can be recognized also behind some other figurative expressions for resistance or undoing, consisting of a verb of destruction and alas, such as (45b) – i.e. the borderline between (A) to (B) is vague.

(45a) ampua alas ‘shoot down (an idea, a plan)’
(45b) repiä alas (lit. tear down) ‘break down (boundaries, conventional values etc.)’

(B) Serialization indicates compositionality. Rudiments of serialized formation occur as niches of a few similar particle verbs, i.e. there is some variation in the verb component, cf. (46a) where yhteen ‘together’ alludes to a confrontation, or (46b) where kiinni ‘shut, fixed, closed’ refers to a state that cannot be changed anymore.

(46a) iskeä yhteen (lit. hit together) ‘clash, lock horns’
    ottaa yhteen (lit. take together) ‘clash, quarrel’
(46b) iskeä kiinni (lit. hit fixed) ‘stabilize (e.g. a dominating position)’
    naulata kiinni (lit. nail fixed) ‘nail down, fix on (e.g. prizes)’
(C) MWEs consisting of verbs and the particles *ilmi* ‘open(ly), apparent(ly)’ and *julki* ‘(in) public, out’ build productive phraseosyntactic patterns, expressing that information is made available or public. In contrast to adverbs like *ulos* ‘out’ or *kiinni* ‘shut, fixed, closed’ which occur both in concrete and in figurative combinations, *ilmi* and *julki* always have a constant abstract meaning, which can explain the stronger serialization. There are both intransitive and transitive series. The kernel verbs are so-called light verbs like *tulla* ‘come’, *antaa* ‘give’, *saada* ‘get’, *tuoda* ‘bring’, but they can be replaced with more specific verbs expressing for example that the publicity was not intended, cf. (47a) vs. (47b). In the transitive pattern, *antaa* ‘give’, *saattaa* ‘put’ or *tuoda* ‘bring’ can be replaced by various speech verbs and their descriptive and expressive variants, cf. (48a–c):

(47a)  *tulla julki* ‘come out, become public’
(47b)  *vuotaa ~ lipsahtaa julki* ‘leak ~ slip out’

(48a)  *tuoda julki* ‘bring into publicity’
(48b)  *lausua ~ puhua ~ sanoa julki* ‘express ~ speak ~ say publicly’
(48c)  *kaakattaa ~ kiljua ~ mölyytellä julki* ‘cackle ~ scream ~ blurt out’

(D) Combinations of verb and directional adverb are often situated on the boundary between regular syntactic constructions and fixed MWEs. At first sight it seems controversial that, according to Kolehmainen (2006: 91, 97, 170), the German separable particle verbs in (49) are lexicalized phraseological (but not idiomatic) units, whereas the corresponding Finnish combinations are not. However, this is not necessarily controversial because the lexicalization strategies in two languages need not be identical. Yet, the difference in the language-specific affinity of such combinations to merge into one lexeme should be proved theoretically. A possible explanation could be related to the grade of semantic-structural autonomy of German and Finnish adpositions and adverbs. Different word order conditions could be relevant, too.

(49)  *weg/ziehen – muuttaa pois* ‘move away’
      *vor/gehen – kulkea edellä* ‘walk ahead’
      *auf/bllicken – katsoa ylös* ‘glance up’
      *hinaus/gehen – mennä ulos* ‘go out’
      *nieder/knien – polvistua alas* ‘kneel down’

In any case it is obvious that lexicalization is mostly combined with semantic specificity. As for the directional adverb *ulos* ‘out’, for instance, the concrete
non-specific meaning is manifest in contexts where the locality inside of something that is left behind is explicated verbally (50a) or when the location is inferable by context and situation, like in (50b), assuming that ‘being in a tunnel’ is already a known fact (contextual ellipsis).

(50a)  _aja u los tunnelista_ ‘drive out of the tunnel, leave the tunnel’
(50b)  _aja u los (Ø)_ ‘leave’

Besides contextual ellipses there are conventionalized ellipses that are not figurative but bear some specific semantic features connected with a certain topic or text type. For example, in reports on road accidents or motor sports _aja u los_ has the conventional meaning ‘drive off the track, swerve off the road’ (51a). The noun _u los+a j o_ (51b) is used particularly in this specific meaning, yet it is difficult to say if it has been derived from the lexicalized phrasal verb. It could as well have been originated as a synthetic compound and then later specialized as a traffic term, of which the specific phrasal verb has been formed analogically, similar to backformation. This makes it difficult to use phrasal input for derivation as a criterion of lexicalizedness of the base, especially as there are synthetic compounds going back to fully transparent non-specific combinations, cf. (52) and (49) above – even if dictionaries codify primarily the idiomatized or specialized compounds and leave the semantically self-evident ones out.

(51a)  _Henkilöauto a j o i u los _sunnuntainen Räyringissä. ‘On Sunday, a passenger car drove off the road in Räyrinki.’
(51b)  _u los+a j o_ (lit. out+driving) ‘driving off the road’ (nomen acti), cf. _U losajo tallentui videolle_. ‘The accident [driving off the road] was videotaped.’

(52)  _muuttaa pois_ ‘move away’ – _pois+muutto_ (N)
      _mennä u los_ ‘go out’ – _ulos+meno_ (N)

Components of lexicalized MWEs cannot be anaphorized. Consequently, if the particle in a combination with a verb is anaphorizable, the combination is free, cf. (53). However, many adverbs lack natural anaphors. For example _kiinni_ ‘fixed, shut, closed’ is not anaphorizable regardless of whether it occurs in concrete or figurative meaning. So, anaphorizability can exclude a combination from phrasal verbs, but lacking anaphorizability cannot be used as evidence of lexicalizedness.

(53)  _Anna _meni _u los._  – _Menikö hän sinne yksin?_ ‘Anna went out. – Did she go there alone?’
Summa summarum: The concept of phrasal verbs deserves to be applied to Finnish, yet, further research is needed to define the limits of the category.

4 Concluding remarks

Compounding is the most common way to form new words in modern Finnish. Prototypical determinative nominal compounds with an underspecified first constituent (N+N) form the most common and still increasing type. Apart from this type many less prototypical compound models are productive, too. Among these, special attention has been paid above to formations showing syntactic features similar to MWEs and/or competing with MWEs. The essential findings can be summarized as follows:

1. In about one third of A+N compounds the adjective agrees in number and case with the head, which does not fulfil the criterion of morphological integrity. However, compound-internal congruence is a recessive feature; there are hardly any neologisms with internal congruence. Compounds with a non-congruent first constituent tend to have a term-like character.

2. Internal inflection also occurs in complex numerals. Numerals with hundreds, thousands etc. are grouped into smaller (still complex) units, thus combining characteristics of non-prototypical compounds and MWEs.

3. In synthetic compounds argument relations of the verb that underlies the head are explicated by case forms, which is a syntactic feature.

4. A prototypical compound cannot be replaced with a phrasal unit of formally identical components. Generally, if such pairs occur, they differ in meaning. Overlap occurs if the modifier is in the genitive, which is the situation for semantically relative adjectives and many deverbal nominalizations. Univerbation strengthens the conceptual unity, and vice versa, conceptualization furthers univerbation.

5. An opposite example of the correlation between conceptual unity and univerbation is represented by Finnish particle verbs. Compound verbs with an adverb or adposition as first constituent are not productive in modern Finnish, partly as consequence of normative language planning. This gap in the system is compensated by “phrasalization”, i.e. keeping apart the components in particle verbs. However, the formation model is far less productive than in English or German, for instance. Apart from singular idioms, serialization, based on phraseosyntactic patterns, occurs in some amount. Drawing the line between lexicalized MWEs and syntactically free combinations requires further research.
6. When the syntactic distributions of a compound and a semantically equal MWE are different, their relation is complementary rather than competing. This applies to simulative compound adjectives/adverbs and corresponding phrasal similes: The latter cannot occur as adjective attributes. Furthermore, while predicative and adverbial simulative compounds can be transformed into phrasal similes, the opposite is not always possible: Only phrasal similes allow expansions in the part that expresses the point of comparison.

The following topics remain for further research: In Finnish, non-figurative MWEs such as fixed collocations and nominations for specific concepts have been so far studied mostly in terminology. In the future, more attention should also be paid to corresponding combinations in standard language. So far, MWEs have been excluded when working out the statistical distribution of different lexeme structure types in the Finnish vocabulary. Another question deserving attention is the role of MWE patterns at the intersection of syntax and lexicon: Besides particle verbs and similes, e.g. light-verb constructions, binomials and serial modification of a specific idiom structure are topics worth of further attention. Several single studies to these areas have been carried out within contrastive phraseology and construction grammar but a systematic overview of MWE patterns is still outstanding.

References

Baldwin, Timothy/Kim, Su Nam (2010): Multiword expressions. In: Indurkhya, Nitin/Damerau, Fred J. (eds.): Handbook of natural language processing. 2nd ed. Boca Raton i.a.: CRC Press. 267–292.

Eronen, Riitta (1996): Monimuotoiset yhdyssanat. In: Kielikello 1996, 4. Internet: http://kielikello.fi.libproxy.helsinki.fi/index.php?mid=2&pid=11&aid=386 (last access: 15.9.2017).

Fleischer, Wolfgang/Barz, Irmhild (2012): Wortbildung der deutschen Gegenwartssprache. 4th ed. Berlin/Boston: De Gruyter.

Häkkinen, Kaisa (1987): Suomen kielen vanhoista ja uusista yhdysverbeistä. In: Sananjalka 29. 7–29.

Häkkinen, Kaisa (1994): Agricolasta nykykielen. Suomen kirjakielien historia. Porvoo: WSOY.

Häkkinen, Kaisa (1997): Kuinka ruotsin kieli on vaikuttanut suomeen? In: Sananjalka 39. 31–53.

Heinonen, Tarja Riitta (2001): Harmaaturkit herkkusuut – bahuvrilihit sanakirjassa ja kielipissa. In: Virittäjä 105. 625–634.

Heinonen, Tarja Riitta (2010): Kuin-vertaukset. In: Virittäjä 114. 348–373.

Hopper, Paul J./Traugott, Elizabeth Closs (2003): Grammaticalization. 2nd ed. (= Cambridge Textbooks in Linguistics). Cambridge, UK: Cambridge University Press.
Hüning, Matthias/Schlücker, Barbara (2015): Multi-word expressions. In: Müller, Peter O. et al. (eds.). 450–467.

ISK (2004) = Hakulinen, Auli et al. (2004): Iso suomen kielioppi. (= Suomalaisen Kirjallisuuden Seuran Toimituksia 950.) Helsinki: SKS.

Jaakola, Minna (1997): Genetiviin kanssa esiintyvien adpositioiden kielipillistumisesta. In: Lehtinen, Tapani/Laitinen, Lea (eds.): Kielipillistuminen. Tapaustutkimuksia suomesta. (= Kieli 12). Helsinki: SKS. 121–156.

Jussila, Raimo (1988): Agricolan sanasto ja nykysuomi. In: Koivusalo, Esko (ed.): Mikael Agricolan kiel. (= Tietolipas 112). Helsinki: SKS. 203–288.

Karlsson, Fred (1983): Suomen kielen äänne- ja muotorakenne. Porvoo: WSOY.

Karlsson, Fred (2004): Finnish (Finno-Ugric). In: Booij, Geert et al. (eds.): Morphology. Morphology. Ein internationales Handbuch zur Flexion und Wortbildung. An international handbook on inflection and word-formation. Vol. 2. (= Handbooks of Linguistics and Communication Science (HSK) 17.2). Berlin/New York: De Gruyter. 1328–1342.

Karlsson, Fred (2015): Finnish. An essential grammar. Translated by Andrew Chesterman. 3rd ed. London/New York: Routledge.

Kielitoimiston sanakirja (2006) = Grönros,Eija-Riitta et al. (2006): Kielitoimiston sanakirja. Osat 1–3. Helsinki: Kotimaisten kielten tutkimuskeskus.

Koivisto, Vesa (2013): Suomen sanojen rakenne. Helsinki: SKS.

Kolehmainen, Leena (2006): Präfix- und Partikelverben im deutsch-finnischen Kontrast. (= Finnische Beiträge zur Germanistik 16). Berlin i. a.: Lang.

Korhonen, Jarno (2018): Fraseologia – Kiinteiden sanayhtymien tutkimus. Helsinki: Finn Lectura.

Lehmann, Christian (1985): Grammaticalization. Synchronic variation and diachronic change. In: Lingua e Stile 20, 3. 303–318.

Mäkisalo, Jukka (2000): Grammar and experimental evidence of finnish compounds. (= Studies in Languages 35). Joensuu: University of Joensuu.

Malmivaara, Terhi (2004): Luupää, puupää, puusilmä. Näkymiä sananmuodostuksen analogisuuteen ja bahuvriihyhydysyssanojen olemukseen. In: Virittäjä 108. 347–363.

Müller, Peter O. et al. (eds.) (2015–2016): Word-formation: An international handbook of the languages of Europe. (= Handbooks of Linguistics and Communication Science (HSK) 40). Berlin/Boston: De Gruyter.

Nenonen, Marja (2002): Idiomit ja leksikko. Lausekeidomien syntaktisia, semanttisia ja morfoloogisia piirteitä suomen kielessä. (= Publications in the Humanities 29). Joensuu: University of Joensuu.

Niemi, Jussi (2009): Compounds in Finnish. In: Lingua & Linguaggio 8. 237–256.

Niinimäki, Anneli (1992): Sanalittojen tiivistyminen yhdysyssanoisiksi. In: Virittäjä 96. 283–286.

Nykysuomen sanakirja (1951–1961) = Sadeniemi, Matti (1951–1961): Nykysuomen sanakirja. Osat I–VI. Porvoo: WSOY.

Öhmann, Emil (1957): Beobachtungen über feste Verbalzusammensetzungen im Finnischen. In: Ural-altaische Jahrbücher 29. 33–37.

Olsen, Susan (2015): Composition. In: Müller, Peter O. et al. (eds.). 364–386.

Pääkkönen, Irmeli (1989): Sanojen äänneasu ja oikeinkirjoitus. In: Vesikansa, Jouko (ed.). 357–382.

Pitkänen-Heikkilä, Kaarina (2016): Finnish. In: Müller, Peter O. et al. (eds.). 3209–3228.

Rahtu, Toini (1984): Suomen nominialkuiset yhdysverbit. In: Virittäjä 88. 409–430.

Räisänen, Alpo (1986): Sananmuodostus ja konteksti. In: Virittäjä 90. 155–163.
Sag, Ivan A. et al. (2002): Multiword expressions: A pain in the neck for NLP. In: Gelbukh, Alexander (ed.): Computational linguistics and intelligent text processing. Third International Conference, CICLing-2002, Mexico City, Mexico, February 17–23, 2002. (= Lecture Notes in Computer Science 2276). Berlin i. a.: Springer. 1–15.

Sajavaara, Paula (1989): Vierassanat. In: Vesikansa, Jouko (ed.). 64–109.

Saukkonen, Paivi (1973): Suomen kielen yhdyssanojen rakenne. In: Commentationes Fenno-Ugricae in honorem Erkki Itkonen. (= Suomalais-Ugrilaisen Seuran toimituksia 150). 332–339.

Schellbach-Kopra, Ingrid (1964): Die Bahuvrihi-Komposita in der alten finnischen Volksdichtung. In: Suomalais-ugrilaisen seuran aikakauskirja – Journal de la Société finno-ougrienne 65. 1–41.

Stein, Stephan (2012): Phraseologie und Wortbildung des Deutschen. Ein Vergleich von Äpfeln mit Birnen? In: Prinz, Michael/Richter-Vapaatalo, Ulrike (eds.): Idiome, Konstruktionen, „verblümte Rede“. Beiträge zur Geschichte der germanistischen Phraseologieforschung. (= Beiträge zur Geschichte der Germanistik 3). Stuttgart: Hirzel. 225–240.

Suomen kielen perussanakirja (1990–1994) = Haarala, Risto (1990–1994): Suomen kielen perussanakirja. Osat 1–3. (= Kotimaisten kielten tutkimuskeskuksen julkaisuja 55). Helsinki: Valtion painatuskeskus.

Toropainen, Tanja (2017): Yhdyssanat ja yhdyssanamaiset rakenteet Mikael Agricolan teoksissa (= Turun yliopiston julkaisuja – Annales Universitatis Turkuensis C 439). Turku: University of Turku. Internet: https://doria.fi/bitstream/handle/10024/143331/AnnalesC%20439Toropainen.pdf?sequence=2 (last access: 25.8.2017).

Tyysteri, Laura (2015): Aamiaiskahvilasta ötökkötarjontaan. Suomen yleiskielen morfosyntaktisten yhdyssanarakenteiden produktiivisuus. (= Turun yliopiston julkaisuja – Annales Universitatis Turkuensis C 408). Turku. Internet: https://doria.fi/bitstream/handle/10024/113113/AnnalesC408Tyysteri.pdf?sequence=2 (last access: 25.8.2017).

Vattinen, Tanja (2003): Vanhan kirjasuomen yhdyssverb. In: Sananjalka 45. 45–66.

Vesikansa, Jouko (1989): Yhdyssanat. In: Vesikansa, Jouko (ed.). 213–258

Vesikansa, Jouko (ed.) (1989): Nykysuomen sanavarat. Porvoo: WSOY.
