

Marja Nenonen

Prototypical Idioms: Evidence from Finnish

Abstract

On the basis of corpus data, it is suggested that a prototypical Finnish idiom is a verb phrase consisting of idiom-prone words, especially basic verbs and body part nouns. The Finnish data indicates that idioms—albeit syntactic phrases—also play an important role in lexicalization, not through the regular lexeme formation processes, i.e. derivation and compounding, but through using idiom constructions with case and/or number inflections, with a relatively simple morphological structure and prototypical simplex words.¹

1. Introduction

Idiom has traditionally been a term that eludes an explicit definition. There are four features, among which there seems to be a consensus among most scholars, although, to some extent, they are all on different conceptual levels (see also Langlotz 2006: 3). First, idioms are **multi-word syntagms**. This feature is a linguistic convention in order to rule out, for example, compounds and monomorphemic words that are also arbitrary by nature (Hockett 1960: 173). Second, idioms are **non-compositional**. Non-compositionality can mean—besides the fact that the meaning of an expression is not decomposable on the basis of the meanings of its parts—also that the form of the expression may not be isomorphic with the referential form (Geeraerts 1995). Third, idioms are regarded as **morpho-syntactically and/or lexically restricted**. However, this feature is always based on post hoc analyses. Finally, in certain aspects idioms are always **conventional** in a sense that they are **institutionalized** (for institutionalization, see also e.g. Brinton and Traugott 2005: 45–47). These conventions may be (relatively) fixed meanings or structural conventions, i.e., construc-

¹ I thank Irma Hyvärinen, Leena Kolehmainen, Jussi Niemi, Esa Penttilä, and an anonymous reviewer for their valuable comments and suggestions on this paper. This study is funded by Academy of Finland, research grant #105825.

tional idioms (Booij 2002, Penttilä 2006). The same conventions must also rule the formation of new idioms: we—both as language speakers and as linguists—must recognize them as idioms and separate them from other, non-idiomatic expressions. (For difficulties in defining the concept of idiom, see also e.g. Wood 1986, Moon 1998, Čermák 2001, Taylor 2002, Langlotz 2006).

According to the criteria above, we can define a **prototypical² idiom**—a multi-word, non-compositional, fixed lexical unit whose meaning is institutionalized (see also Häkkinen 2000: 8–10). Prototypical units of language have been studied mostly with respect to lexicon and grammatical categories (e.g. Rosch 1978, Taylor 1998, 2003). We may, however, also take advantage of the prototype theory when studying larger linguistic units such as phrasal idioms.³ Not every idiom can be called prototypical, although it fulfills the conditions mentioned above, because a prototype usually refers to a basic level category. In order to exploit the prototype theory, we must find the conceptual core of the category of “idiom.” In this respect we have to turn to the data: it seems that there are certain kinds of words and structures that are typical of idioms, as we will see below in the Finnish data.

1.1 A brief overview to the Finnish lexicon

As regards lexicon, Finnish language is fairly transparent—at least phonologically if not always semantically—and it is relatively easy to separate the stems from the endings. Moreover, Finnish morphology is extremely rich: nouns have 15 cases and tens of derivative affixes, not to mention compounding that is highly productive as well. And verb morphology is even richer (see e.g. Karlsson 1983: 356–357, Hakulinen et al. 2004: 106–422).

² A prototype can be understood in at least three ways (Taylor 2003: 63–64): a) as a specific instance of a category; b) as a specific kind of entity (on this approach, a particular entity instantiates the prototype); or c) as an even more abstract notion, which captures the conceptual ‘centre’ of the category. In this study, the concept of prototype is placed somewhere between b and c.

³ Prototypicality and idiomaticity have been studied by Akimoto (1992), who suggests that in verb phrase idioms complement nouns lose their prototypical meaning and even their prototypical nouniness. In a more recent paper, Akimoto (1994) discusses typical idioms in certain languages and regards body part noun idioms as universals, mostly occurring in the V+O pattern, a notion that is also the gist of this study.

Figure 1 presents a schematic overview of Finnish lexical units, based on the ratio of the type frequencies of basic words, derivations and compounds calculated from the largest Finnish dictionary, *Nykysuomen sanakirja*, which contains ca. 201,000 entries (according to Niemikorpi (1991: 154), the ratio for basic words, derivations and compounds is 8.6 : 26.6 : 64.8, respectively). If the ratio were sketched on the basis of token frequencies, the size of the circles in Figure 1 would naturally be the opposite, since basic words are the most frequently used in a language, whereas the textual frequency of compounds is usually relatively low. The two-sided arrow at the bottom represents the morphological complexity of lexical items, from monomorphemic words to structurally complex phrasal units.

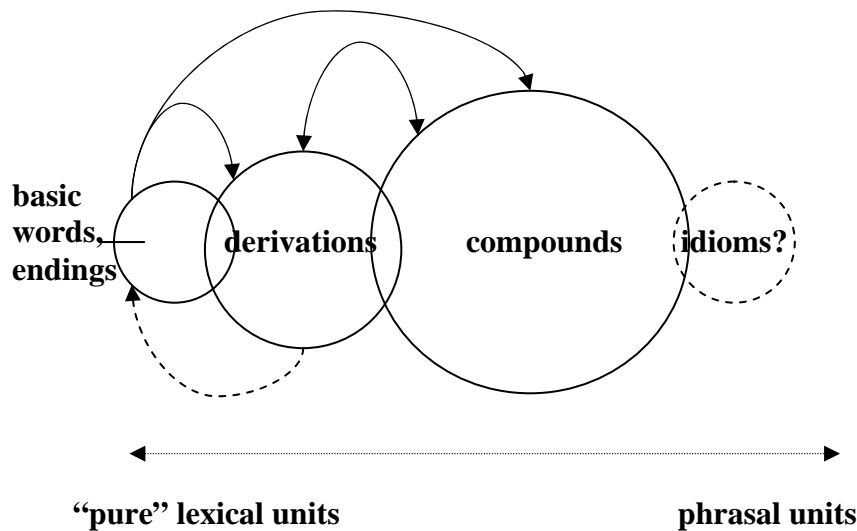


Figure 1. The structure of Finnish lexicon in a nutshell.

Traditionally, in addition to single words, also the derivative and inflectional endings are located in the lexicon, which serves our purposes in describing the current Finnish lexical system. Naturally, the modules in Figure 1 overlap, since, for example, most of the derivations derive from basic words, while some of the opaque derivations have become basic words. Compounds, the largest group in the Finnish lexicon, may contain basic words and derivations as well, and many compounds are also derivable. The “idioms” circle is marked by dotted lines, since, so far, we neither know the size nor the structure of the idiom proportion accurately.

In the following sections, our aim is to look deeper into the lexical and morphological structure of Finnish idioms.

1.2 The Finnish data

The Finnish data presented here were originally collected for a PhD study (Nenonen 2002) on Finnish verbal idioms.⁴ The data are collected from various sources, mainly from three different text corpora:

a) **The literary corpus**, a sample of written Finnish fiction (10 juvenile books, approximately 100,000 running words, for references, see Nenonen 2002: 138) was read through and around 3,000 verbal idioms and noun phrase idioms were manually collected, in order to collect a relevant sample of relative frequencies and typical vocabulary patterns of Finnish idioms. The juvenile books were chosen as test material because they were assumed to consist plenty of colloquial expressions and therefore idiomatic material.

b) **The Karjalainen Corpus**, *Karjalaisen korpus*, a 34.5-million-word token computer-based newspaper corpus of Finnish, based on the newspaper *Karjalainen* (vols. 1991–1998, Joensuu, Finland), was used in collecting frequencies and meanings of the individual idiom-prone words. The typical idioms collected from the literary corpus were studied more closely by means of the *Karjalainen corpus*, using corpus and database tools (Laine and Virtanen 1999, designed for collecting frequencies; Virtanen and Pajunen 2000, designed for concordance and collocation analyses). Also the most idiom-prone words were separated and investigated more closely.

c) **The Finnish Language Bank**, *Kielipankki*, (<http://www.csc.fi/kielipankki>), contains over 100 million words, including also the *Karjalainen Corpus*. The other large Finnish text corpora of *Kielipankki* have also been used for help in finding examples of particular idioms and idiom-prone words.

d) Moreover, several Finnish **dictionaries** have been instrumental, especially in collecting body part noun idioms and studying the different meanings of basic verbs (for detailed information, see Nenonen 2002: 137).

⁴ Verbal idioms cover both verb phrase idioms, i.e. a verb + complement(s), and idioms that consist of a subject + a predicate + an open place for a complement, see also fn. 5 below.

2. Prototypical Finnish idioms

On the basis of a sub-corpus of 3,066 idiom tokens manually collected from the literary corpus, it seems that the most typical Finnish idiom is a **verb phrase idiom** that usually consists of a finite verb (or an infinitive) and one or more complements,⁵ as in (1). Another typical group, **noun phrase idioms**, does not necessarily include a specific verb. Typically, a noun phrase idiom complements a verb phrase, like (2), which is usually realized as a predicative with a copula, or like (3), which may appear with various verbs.

- (1) *potkaista tyhjä-ä*
kick empty-PTV
'to die,' a common Finnish idiom for *kick the bucket*
- (2) *helppo nakki*
easy wienie
'an easy thing to do, a piece of cake'
- (3) *naki-t silm-i-llä*
wienie-PL eye-PL-ADE
'heavily drunk'

In addition to these typical groups, there are also “other idioms,” which do not fit either of these classes, e.g. sentence-like idioms and multi-word interjections. Table 1 presents the distribution of types and tokens of different type of idioms in the literary corpus.

⁵ Sometimes also a subject and a verb with an open place for a complement is possible (Vilkuna 1989: 157–176; for some other languages, see e.g. Reichstein 1973, Nunberg et al. 1994: 525, Fleischer 1997: 99–103, O’Grady 1998: 287–299). However, the typical word order in these idioms is COMP + V + N, e.g. *Minu-lta menee hermo-t*, I-ABL go-3rd nerve-PL, literally ‘from me go (disappear) the nerves,’ ‘I’m losing my nerves,’ where ‘losing one’s nerves’ is the actual idiom, and ‘nerves’ is the subject of a clause. Often this idiom type belongs to an experiencer construction (see also examples (9–10) below).

	verbal idioms	noun phrase idioms	other idioms	total
types	1282	492	345	2119
tokens	1973	705	388	3066

Table 1. Idioms collected from the literary corpus (100,000 running words, Nenonen 2002: 54).

Since the corpus is highly restricted, most of the idioms appear only once, as we see in Table 1. This is not surprising—in general, the textual frequency of idioms is rather low (Moon 1998: 60, Biber et al. 1999: 989). In fact, only six idioms and five constructional idioms appeared in the literary corpus ten times or more (Nenonen 2002: 55; for constructional idioms, see Chapter 2.4. below). However, even the present small sample of corpus data indicates that some words (and also some grammatical categories) are especially **idiom-prone** (for corresponding English data, see Moon 1998: 75–87). The next section will offer a closer look at these words.

2.1 Idiom-prone words

The term idiom-prone here is adapted from Taverner's (1977) definition: idiom-prone words are **productive** in a sense that they appear in several idioms, **highly frequent** in large text corpora, and, moreover, they are **polysemous**. These words include the **basic verbs and body-part nouns** that will be discussed in detail below.

The most common verbs are all so-called **basic** or **nuclear verbs** (Newman 1996, Stubbs 1986). According to Stubbs (1986: 105), nuclear words are pragmatically neutral; they are less specialized in meaning and can thus occur in a wide range of contexts and collocations. It is noteworthy that the aspects that are characteristic of basic words, i.e. generality of meaning, frequency of use and simplicity of form, bring basic words closer to function words (for differences of functional and lexical units in language, see e.g. Biber et al. 1999: 55).

On the one hand, basic verbs are clearly lexical words: they have a lexical meaning, they show morphological variation, they are heads of phrases, they belong to an open class, and their number (as verbs in

general) is large. On the other hand, basic verbs are close to grammatical words in a sense that they are frequent and short, their number is small compared to verbs in general, and they are lexically rather empty *per se*, because the prevailing meaning is context-dependent. This means that basic verbs may be grammaticalized, which increases the generality of meaning; in the purest fashion, they represent only the basic semantic features of their domains (Bybee et al. 1994: 9).

In addition to verbs, some nouns prove to be idiom-prone as well. Especially the most common **body part nouns**—which refer to the most prototypical body parts—tend to appear in idioms: according to the corpora and Finnish dictionaries, *pää* ‘head,’ *silmä* ‘eye,’ and *käsi* ‘hand’ are the most common body part names in idioms (see also Akimoto 1994, for the same phenomenon in Japanese, English, French and German, Nenonen 2002, for Finnish, Niemi 2004, for Swedish, and Mulli 2007, for German).

Not unexpectedly, a typical Finnish phrasal idiom is a verb phrase that consists of a basic verb and an inflected noun, as in examples (4–6). The examples below are chosen so that every complement is in the illative case but the meaning of the phrase differs, in order to show the idiomatic nature of the phrases. In (4) and (5) the syntactic structure is identical, ‘pull’ followed by a body part noun in the illative case, but the meanings are quite different, since the direction of the verb ‘pull’ is different in each example (for Finnish ‘pull+comp’ constructions, see also Niemi 2007). Example (6), on the other hand, is part of an experiencer construction, which is typical for the idiomatic, intransitive use of the verb *ottaa*, ‘take’ (see also examples (9)–(10), below).

- | | | |
|-----|--------------|-------------------------|
| (4) | <i>vetää</i> | <i>nenä-än/nen-i-in</i> |
| | pull | nose-ILL/nose-PL-ILL |
| | | ‘beat someone’ |
| (5) | <i>vetää</i> | <i>naama-an</i> |
| | pull | face-ILL |
| | | ‘eat fast and greedily’ |
| (6) | <i>ottaa</i> | <i>pää-hän</i> |
| | take | head-ILL |
| | | ‘annoy’ |

Table 2 presents the frequencies of the 20 most frequent verbs and nouns of the verbal idioms collected from the literary corpus (Table 1). It is

noteworthy that the ten most idiomatic verbs cover 50% of all the verbs in the verbal idiom types of the sample. These verbs also belong to the most frequent verbs in Finnish in general.

Verb	Types	Tokens	Freq	Noun	Types	Tokens	Freq
<i>olla</i> ‘be’	264	404	1	<i>silmä</i> ‘eye’	27	35	304
<i>ottaa</i> ‘take’	67	142	44	<i>mieli</i> ‘mind’	23	55	91
<i>saada</i> ‘get’	54	87	10	<i>pää</i> ‘head’	23	51	291
<i>mennä</i> ‘go’	51	78	68	<i>suu</i> ‘mouth’	14	22	763
<i>pitää</i> ‘keep’	42	89	29	<i>naama</i> ‘face’	13	18	9466
<i>vetää</i> ‘drag’	41	73	614	<i>asia</i> ‘thing’	11	14	41
<i>tulla</i> ‘come’	37	65	16	<i>korva</i> ‘ear’	11	18	1225
<i>tehdä</i> ‘do’	35	82	30	<i>aika</i> ‘time’	10	14	20
<i>käydä</i> ‘fit’	32	46	87	<i>sana</i> ‘word’	9	12	114
<i>panna</i> ‘put’	50%	27	361	<i>turpa</i> ‘trap’ (‘mouth’)	9	14	7844
<i>lähteä</i> ‘leave’	26	45	106	<i>nauru</i> ‘laughter’	8	12	3100
<i>antaa</i> ‘give’	22	66	28	<i>onni</i> ‘happiness’	8	10	809
<i>pistää</i> ‘stick’	21	26	1099	<i>henki</i> ‘spirit’	7	18	377
<i>heittää</i> ‘throw’	19	29	754	<i>kuvio</i> ‘figure’	7	9	1849
<i>päästä</i> ‘get’	15	29	106	<i>käsi</i> ‘hand’	7	10	168
<i>jäää</i> ‘stay’	12	17	90	<i>matka</i> ‘trip’	7	9	248
<i>katsoa</i> ‘look’	9	19	98	<i>niska</i> ‘neck’	7	10	2262
<i>lyödä</i> ‘hit’	8	10	763	<i>puhe</i> ‘speech’	7	11	444
<i>nähdä</i> ‘see’	8	9	69	<i>aivot</i> ‘brains’	6	21	3282
<i>painaa</i> ‘press’	7	9	614	<i>jalka</i> ‘leg’	6	6	423

Table 2. 20 most common verbs and nouns in different verbal idioms (there were altogether 327 different verbs and 548 different nouns collected from the idioms of the literary corpus, Nenonen 2002: 57). *Freq* = frequency ranking of *the Finnish frequency dictionary*⁶ (Saukkonen et al. 1979).

The arbitrary nature of idioms becomes clearly visible, if we try to generate idioms on the basis of the list in Table 2. Any combination whatsoever is not possible: for example, combinations *ottaa silmä-än* ‘take eye-ILL’ ‘distract, strike one’s eye,’ *ottaa pää-hän* ‘take head-ILL’ ‘annoy,’ or *ottaa*

⁶ *The Finnish frequency dictionary* contains Finnish nuclear vocabulary, 43,670 different words that cover 90% of a c. 400,000-word corpus (Saukkonen et al. 1979: 7).

su-i-hin ‘take mouth-PL-ILL’ ‘have oral sex,’ are idioms, but *ottaa miele-en* ‘take mind-ILL’ and *ottaa naama-an* ‘take face-ILL’ are not.

The following case studies will provide an overview of two of the most idiom-prone words, the verb *ottaa* ‘take’ and the noun *silmä* ‘eye,’ using the *Karjalainen corpus* and dictionaries. The purpose is to map the general usage of these words. At the same time, the aim is to locate idioms in the lexicon.

2.2 The most typical verb: ‘take’

On the grounds of the data, we may claim that in addition to the maximally empty verb *olla* ‘to be,’ *ottaa* ‘take’ is the most idiom-prone verb in Finnish. This frequently used verb has several dictionary meanings and belongs to many constructions that vary in their degree of idiomaticity (see also Ruhl 1999, for corresponding use of English *take*, and Newman 1996, for the verbs ‘give’ and ‘take’ in various languages).

The usage of the verb *ottaa* ‘take’ has been observed using the *Karjalainen corpus* and *Nykysuomen sanakirja*, which reports no fewer than 118 different meanings for this entry (Jussila 1988: 91). On the basis of corpus studies, it appears that the verb *ottaa* is most commonly used with abstract meanings (often metaphorically or in different constructions), whereas the concrete meaning, i.e. ‘picking’ or ‘grasping,’ is relatively rare. For an overview of the use of the verb, Table 3 presents the division of the meanings in a random sample of 500 concordances of the verb *ottaa* in the *Karjalainen corpus* and in the literary corpus. The division of meanings is based on the classification of *Nykysuomen sanakirja* (for more details, see Nenonen 2002: 100–103).

Meaning of <i>ottaa</i>	Karjalainen corpus	Literary corpus
‘pick’	5	34
‘remove’	32	58
abstract, specialized meanings	412	312
intransitive use	5	66
fixed phrases	46	30
Total	500	500

Table 3. Distribution of the meanings of the verb *ottaa* in a sample of 500 tokens from the *Karjalainen corpus* and the literary corpus.

In Table 3 we see that the verb *ottaa* is used most frequently in “abstract or specialized meanings” in the *Karjalainen corpus* (newspaper text). In the literary corpus, which partly imitates spoken language, the distribution differs a little to the advantage of concrete meanings. The meanings vary from rather concrete expressions for dressing, eating, possessing, moving, etc. (e.g. *ottaa takki* ‘take coat, i.e. put on a coat,’ *ottaa makkaraa* ‘take sausage-PTV, i.e. eat some sausage,’ *ottaa lääke* ‘take drug, i.e. have a dose of drug,’ *ottaa haltuunsa* ‘take into one’s possession,’ *ottaa mukaan* ‘take with’) to more abstract expressions for obligations, choosing, regarding, understanding, etc. (e.g. *ottaa huomioon* ‘take into account, pay attention to,’ *ottaa aikaa tehdä jtk* ‘it takes time to do sth,’ *ottaa raskaasti*⁷ ‘take sth hard’). Especially the latter examples could also be placed among “fixed phrases,” which is one of the classes included in the dictionary. Fixed phrases, such as (7–8), is a class that can also be called idioms (see also examples 6 above and 16, 19 below).

- (7) *ottaa kanta-a*
 take attitude-PTV
 ‘speak out’

⁷ *Ottaa ADJ-sti*, ‘take ADJ-adverb derivative ending,’ is also a constructional idiom, that may be filled up with suitable adjectives, like *kevyesti* ‘light-DER,’ or *paha-sti* ‘bad-DER.’

- (8) *ottaa ylös*
 take up-to
 ‘take down, take notes’

In this study, most idioms fall into the largest group, the class of abstract or specialized meanings. It is obvious that the verb *ottaa* has a strong tendency to combine semantically with other words, and some of these combinations are institutionalized, which is a step towards lexicalization. Some of these combinations may also be regarded as phrasal verbs, like (7) and (8) (see also Kolehmainen 2006).

Intransitive use of the verb *ottaa* is typically more common in informal, colloquial texts, especially in experiencer constructions that consist of an experiencer in the partitive case and a complement in a locative case, as examples (9) and (10) indicate. Usually the complement is a body part noun, which is typical of idioms, as we have seen before.

- (9) *Mummo-a ottaa sydäme-stä.*
 granny-PTV takes heart-ELA
 ‘Granny is having a heart attack.’
- (10) *Minu-a ottaa aivo-on/pää-hän/kallo-on.*
 I-PTV takes brain-ILL/head-ILL/skull-ILL
 ‘I am irritated.’

2.3 The most typical noun: ‘eye’

A closer look at the meaning(s) of the most idiom-prone noun *silmä*, ‘eye,’ reveals interesting features of this particular noun, comparable to the verb *ottaa* ‘take’ above. Body part nouns are an interesting area in linguistics, since they tend to be an issue of not only lexicalization but grammaticalization as well. They seem to have a high facility for abstraction; for example, Deignan and Potter (2004) have reported a strong tendency of body nouns to be used non-literally in English and Italian. In their English and Italian data, around 50% of the citations of *eye(s)* were non-literal (Deignan and Potter 2004: 1236). In Finnish, the same tendency is even stronger. As shown below, less than half of the occurrences of noun *silmä* are used in its concrete meaning ‘an organ of sight.’ Otherwise it appears in a more or less abstract sense—also in metaphors and in idioms, like (11–12).

- (11) *iskeä silmä-ä*
hit eye-PTV
'wink at sb'
- (12) *pistää silmä-än*
stick eye-ILL
'stick out'

When we take a closer look at morphology of the noun *silmä*, we see that some cases are more idiomatic than others.⁸ Figure 2 presents the distribution of the ten most common case/number inflected forms of the noun *silmä*. Every case is analyzed by taking a random sample of 100 items per inflected form from the *Finnish Language Bank*. According to the *Karjalainen corpus* (Laine and Virtanen 1999), the lemma frequency of the noun *silmä* is 5,882, i.e. 173/one million words. The summed frequency of the ten most frequent cases, presented in Figure 2, is 4,359, i.e. 74% of all occurrences of the noun *silmä* 'eye' in the corpus.

⁸ Generally, some cases are more idiom-prone than others in Finnish. In Nenonen (2002: 59) it is shown that there are some highly idiom-prone cases especially in verb phrase idioms, like the partitive, the illative, and the instructive, which are more frequently used in idioms than in Finnish in general. On the other hand, the second most frequent case in Finnish, the genitive, is rare in idioms.

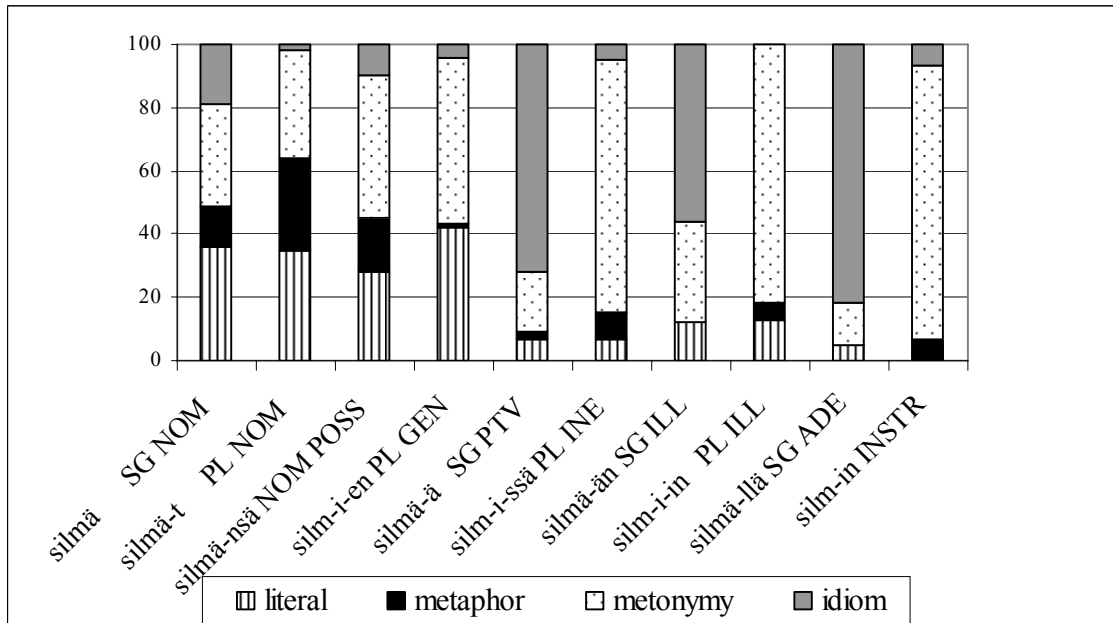


Figure 2. The proportion of literal vs. figurative (i.e. metaphorical, metonymic, idiomatic) meanings of the ten most common inflected forms of the noun *silmä* ‘eye’ in the *Finnish Language Bank* (sample of 100 occurrences/case).

In Figure 2, we notice clearly that there are three cases—partitive, illative, and adessive—that are extremely common in idioms of this sample. In addition to the nominative case, partitive and illative (see examples 11–12 above) are two of the most frequent cases in Finnish idioms (Nenonen 2002: 59). The high proportion of idioms in some cases can also be explained by high textual frequency of certain idioms, like the adessive case in (13), where 47 of the 100 tokens were included in that particular idiom.

- (13) *pitää silmä-llä*
 keep eye-ADE
 ‘keep an eye on sb’

As can be seen in Figure 2, over 80% of the inflected forms in the sample have a non-concrete meaning. Only the typical grammatical cases, the nominative singular, the nominative plural and the genitive plural, are more frequently used with a literal meaning (they also happen to be the most frequent cases in Finnish in general). On the other hand, the instructive

case that is very rare in modern Finnish, only appears with non-literal meanings, like *omin silmin* and *kirkkain silmin* in (14–15).

- (14) *omi-n* *silmi-n*
 own-INSTR eye-INSTR
 ‘with one’s own eyes’

- (15) *kirkkai-n* *silmi-n*
 bright-INSTR eye-INSTR
 ‘bright-eyed, pretending not to lie’

The instructive case in examples (14–15) refers to typical abstract usage of the noun ‘eye,’ i.e. metonymic ‘seeing’ or ‘looking’ (see also Deignan and Potter 2004).

2.4 Plural: a grammatical index of idiom

In addition to words, some grammatical categories can be idiom-prone as well, for example, the plural in Finnish. In idiomatic phrases, the plural ending is not used to refer to numerous entities but rather to a single, recurrent event as in (16).

- (16) *ottaa* *pitkä-t/lähdö-t/hatka-t*
 take long-PL/leaving-PL/hatka⁹-PL
 ‘leave, quit, take off, take a hike’

The idioms in (16) belong to the idiom family (Nunberg et al. 1994: 504) or the construction *ottaa N:t* ‘take N-PL,’ which allows various plural nouns (or even adjectives such as ‘long’) to be added to the construction in order to create an idiom. In general, a basic verb and a noun in the nominative plural constitute a typical constructional idiom in Finnish. A **constructional idiom**, according to Booij (2005: 83), is “a fixed syntactic pattern in which some positions may be filled by all kinds of words of the right category, whereas other positions are filled by specific morphemes or words.” Another example of the constructional idiom is an eponymous verb phrase *tehdä N_{prop}:t* ‘do N_{prop}-PL’ (17), in which you may add any proper

⁹ *Hatkat* is a typical idiomatic isolate, sometimes also referred as a cranberry word, i.e. it is a unique word that appears only in a certain idiom (see also Nenonen and Niemi 1999, Nenonen 2007).

name (or sometimes even a common noun) after the verb in order to form an idiom.

- (17) *tehdä väyryse-t*
do Väyrynen-PL
e.g. ‘sleep on it,’ ‘change one’s attitude,’ ‘cite oneself’

This idiom construction is highly context-dependent, as we can see in (17), where the interpretations are picked up among various examples on the Internet (Paavo Väyrynen is a well-known Finnish politician). The orthography of the idiom varies, but usually also the proper name begins with a small letter—this is a marker of idiomaticity as well.¹⁰ Generally, the interpretation of the idiom derives from a certain salient characteristic or action of the one the name refers to (see also Karlsson 2000).

Another example of the constructional idiom is a combination of a basic verb and a mass noun plural, as in (18), or a body part noun, as in (19). In (18), the idiom refers to a certain situation, e.g. a break from work, not to the “stuff” that we drink; one might as well have a nice cup of tea instead of coffee, and the drink is usually consumed with pastry (Niemi, Nenonen and Penttilä 1998: 296). On the other hand, the idiom in (19) refers to a situation that includes heavy and uncontrolled drinking.

- (18) *juoda kahvi-t*
drink coffee-PL
‘have some coffee’

- (19) *ottaa/vetää/juoda* *persee-t* (*ola-lle*)
take/drag/drink ass-PL (shoulder-ALL)
‘get heavily drunk’

In all these examples (16–19), the connective aspect is not the plurality itself, but certain recurrency and abstractness. In this sense, we may regard unpredictable number as an indexical marker of idiomaticity in Finnish (Niemi et al. 1998, Nenonen 2002).

¹⁰ Cf. the corresponding English structure, *do a N_{prop}*, e.g. *do a Chomsky*, where the proper noun is preceded with an indefinite article (Penttilä 2006). As in Finnish, also in English the prevailing meaning of the idiom is mostly dependent on the context (for idioms and constructions in general, see also Goldberg 1995, Jackendoff 1997, Booij 2002).

3. Discussion

On the basis of Finnish data, it seems that prototypical idioms are morphologically relatively simple. They consist of short, frequent basic verbs and inflected monomorphemic nouns. The most common idiom words refer to the basic level categories and belong to the most frequently used words in Finnish. In this respect, we may say that prototypical idioms are made of prototypical elements of language. In Table 4 we see that basic verbs and body part nouns form the majority of the most common idiom words (the numbers are based on data in Table 2 above). It must also be noted that the ten most common verbs in idioms—all basic verbs—already cover over a half of all the verbs that occur in idioms in the present corpus.

	Verbs	Nouns
Basic verbs and body part nouns among all idiom words in verbal idioms (<i>n</i> of tokens = 1973)	58%	23%
Basic verbs and body part nouns among the 20 most common words in verbal idioms (<i>n</i> = 1361)	93%	70%
Basic verbs and body part nouns among the 10 most common words in verbal idioms (<i>n</i> = 1102)	100%	80%

Table 4. The proportion of basic verbs and body part nouns in the verbal idioms shown in Table 1.

All in all, the Finnish data indicate that idioms participate in lexicalization, not through the regular lexeme forming processes, i.e. derivation and compounding, but through idiom constructions with particular case and/or number inflections. In Figure 3, the implication arrows illustrate the dynamics of idiom formation in the Finnish lexicon. Above, we saw that derivations and compounds mainly derive from basic words and that compounds may include derivations as well, and vice versa (Figure 1). In Figure 3, the focus is on idioms. The figure is simple: basic words and inflectional endings are the main building blocks of Finnish idioms. Derivations are extremely rare in phrasal idioms, and so are also compounds (the latter, however, may sometimes appear in idioms).

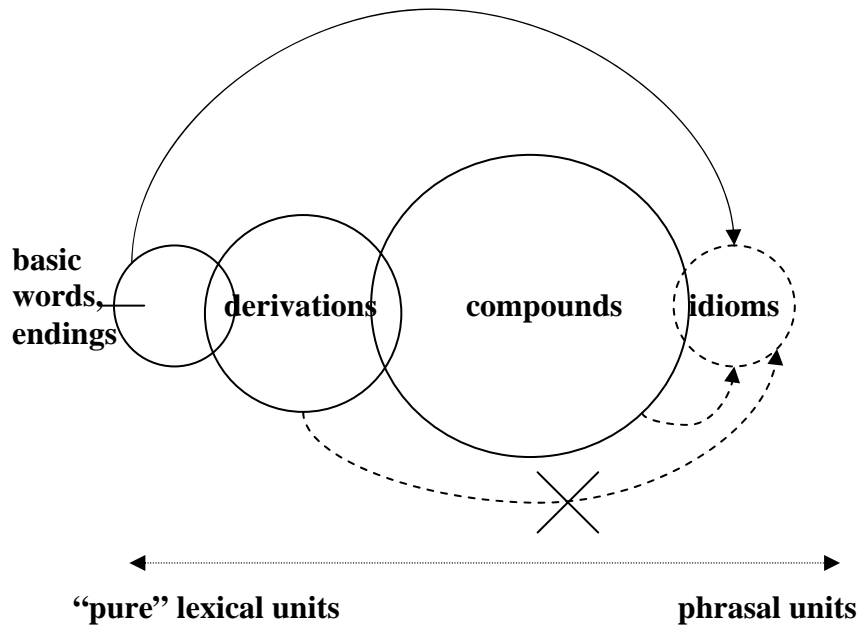


Figure 3. Idioms with relation to the Finnish lexicon.

What then is the main reason for the fact that derivations and compounds so reluctantly compose idioms? There may be several reasons, but the most obvious one is the simplicity of form: phrasal idioms are complex enough without derivations or compounds. In this respect, we may regard idiomatization as the “third party” of lexeme formation; it participates in lexeme formation through inflection in addition to the traditional methods of morphological word formation, i.e. derivation and compounding (see also Hyvärinen 1992: 36, Booij 2002).

Prototypical idioms are also close to single lexical units in the sense that their component words lose much of their semantics. The basic verbs are close to grammatical or functional words, and so are also many of the complements, although the range of complement nouns is wider. However, at least body part nouns tend to lose their prototypical nouniness, and thus make the phrases more idiomatic with syntactic irregularities and semantic discrepancy from their original meaning (Akimoto 1992: 236). The same tendency can also be found in verbs, for example, in the intransitive use of the typically transitive verb *ottaa*, ‘take,’ but, as we saw in section 2.2 above, this use may be rather marginal—although eye-catching.

To summarize, **prototypical idioms** are verb phrases, consist of basic words, and are morphologically simple. On the other hand, **prototypical words** that idioms are made of belong to basic-level categories, appear

very frequently in text corpora, and are morphologically simple, as well. What differentiates these idioms from any prototypical phrases? Nothing, in principle, if we look only at the morphological structure of the phrasal units.¹¹ As suggested in every idiom study: it is the non-compositionality and unpredictability of meaning that makes the difference.

List of abbreviations

ABL	ablative
ADE	adessive
ALL	allative
DER	derivational suffix
ELA	elative
GEN	genitive
ILL	illative
INE	inessive
INSTR	instructive
PL	plural
POSS	possessive
PTV	partitive
SG	singular

References

- Akimoto, Minoji (1992) Prototypicality and idiomaticity. In *The 18th LACUS Forum 1991*. Lake Bluff, IL: Linguistic Association of Canada and the United States, pp. 235–245.
- Akimoto, Minoji (1994) A typological approach to idiomaticity. In *The 20th LACUS Forum 1993*. Lake Bluff, IL: Linguistic Association of Canada and the United States, pp. 459–467.
- Biber, Douglas, Johansson, Stig, Leech, Geoffrey, Conrad, Susan & Finegan, Edward (1999) *Longman Grammar of Spoken and Written English*. London: Longman.

¹¹ There is also psycholinguistic evidence that at least verb phrase idioms are syntactically fully parsed during comprehension like any non-idiomatic phrase, although in general, idioms may be easier to process (e.g. Flores D’Arcais 1993, Nenonen et al. 2002, Vainio and Nenonen 2007). The meaning of an idiom, however, may be stored in the mental lexicon as a holistic unit. There is also evidence that especially noun phrase idioms may be stored as wholes in the mental lexicon, in the same way as compounds (Nenonen et al. 2002).

- Booij, Geert E. (2002) Constructional idioms, morphology, and the Dutch lexicon. *Journal of Germanic Linguistics* 14: 301–329.
- Booij, Geert E. (2005) *The Grammar of Words: An Introduction to Linguistic Morphology*. Oxford: Oxford University Press.
- Brinton, Laurel J. & Traugott, Elizabeth Closs (2005) *Lexicalization and Language Change*. Cambridge: Cambridge University Press.
- Bybee, Joan, Perkins, Revere & Pagliuca, William (1994) *The Evolution of Grammar: Tense, Aspect, and Modality in the Languages of the World*. Chicago: University of Chicago Press.
- Čermák, František (2001) Substance of idioms: Perennial problems, lack of data or theory? *International Journal of Lexicography* 14: 1–20.
- Deignan, Alice & Potter, Liz (2004) A corpus study of metaphors and metonyms in English and Italian. *Journal of Pragmatics* 36: 1231–1252.
- Fleischer, Wolfgang (1997) *Phraseologie der deutschen Gegenwartssprache*. Tübingen: Niemeyer.
- Flores d'Arcais, Giovanni B. (1993) The comprehension and semantic interpretation of idioms. In Cristina Cacciari & Patrizia Tabossi (eds.), *Idioms: Processing, Structure, and Interpretation*, pp. 79–98. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Geeraerts, Dirk (1995) Specialization and reinterpretation in idioms. In Martin Everaert, Erik-Jan Van Der Linden, André Schenk & Rob Schreuder (eds.), *Idioms: Structural and Psychological Perspectives*, pp. 57–74. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Goldberg, Adele E. (1995) *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press.
- Häkkinen, Kaisa (2000) Idiomien määritelmää [Definitions of an idiom]. In Eeva Lähdemäki & Yvonne Bertills (eds), pp. 1–15. Åbo: Åbo Akademis tryckeri.
- Hakulinen, Auli, Vilkuna, Maria, Korhonen, Riitta, Koivisto, Vesa, Heinonen Tarja Riitta & Alho, Irja (2004) *Iso suomen kielioppi* [The big Finnish grammar]. Helsinki: SKS.
- Hockett, Charles F. (1960) *A Course in Modern Linguistics*. Third edition. New York: Macmillan.
- Hyvärinen, Irma (1992) Zum Festigkeitsgrad des Verbs und zur aktionalen Reihenbildung bei deutschen und finnischen Verbidiomen: Ein Werkstattbericht im Rahmen des Projekts “Kontrastive Verbidiomatik Deutsch-Finnisch”. In Jarmo Korhonen (ed.), *Phraseologie und Wortbildung—Aspekte der Lexikonerweiterung*, pp. 21–38. Linguistische Arbeiten 284, Tübingen: Max Niemeyer Verlag.
- Jackendoff, Ray (1997) *The Architecture of the Language Faculty*. Cambridge: The MIT Press.
- Jussila, Raimo (1988) Leksikografien luokittelu kvantitatiivisesta näkökulmasta [Lexicographic classification from quantitative point of view]. In Heikki Nyysönen & Olli Kuure (eds.), *XV kielitieteen päivät Oulussa 13.–14.5.1988*. Acta Universitatis Ouluensis B 14, pp. 85–92. Oulu: University of Oulu.
- Karjalaisen korpus* (*The Karjalainen corpus*), a 34.5-million-word token computer-based newspaper corpus of Finnish based on the newspaper Karjalainen (vols.

- 1991–1998, Joensuu, Finland). Availability: Kielipankki (Language Bank of Finland), <http://www.csc.fi/kielipankki/>.
- Karlsson, Fred (1983) *Suomen kielen äänne- ja muotorakenne* [Finnish phonology and morphology]. Porvoo: WSOY.
- Karlsson, Göran (2000) Lehtiväen idiomissa erikoinen monikon käyttötapa [A peculiar use of the plural in the journalistic idiom]. In Eeva Lähdemäki & Yvonne Bertills (eds), pp. 96–104. Åbo: Åbo Akademis Tryckeri.
- Kolehmainen, Leena (2006) *Präfix- und Partikelverben im deutsch-finnischen Kontrast*. Finnische Beiträge zur Germanistik, Vol. 16. Frankfurt am Main: Peter Lang.
- Laine, Matti & Virtanen, Patrik (1999) *WordMill Lexical Search Program*. Center for Cognitive Neuroscience, University of Turku, Finland.
- Langlotz, Andreas (2006) *Idiomatic Creativity: A Cognitive-Linguistic Model of Idiom-Representation and Idiom-Variation in English*. Amsterdam: John Benjamins.
- Lähdemäki, Eeva and Bertills, Yvonne (eds.) (2000) *Fennistica Fausta in honorem Mauno Koski septuagenarii*. Åbo: Åbo Akademis tryckeri.
- Moon, Rosamund (1998) *Fixed Expressions and Idioms in English: A Corpus-Based Approach*. Oxford: Clarendon Press.
- Mulli, Juha (2007) Lexical observations on German idioms. In Marja Nenonen & Sinikka Niemi (eds.), pp. 193–199. Joensuu: University of Joensuu.
- Nenonen, Marja (2002) *Idiomit ja leksikko: Suomen kielen lausekeidiomien syntaktisia, semanttisia ja morfologisia piirteitä*. [Idioms and the lexicon. Syntactic, semantic, and morphological features of Finnish phrasal idioms]. Doctoral dissertation. University of Joensuu Publications of the Humanities 29. Joensuu: University of Joensuu.
- Nenonen, Marja (2007) Unique but not cranberries: Idiomatic isolates in Finnish. In Marja Nenonen & Sinikka Niemi (eds.), pp. 213–223. Joensuu: University of Joensuu.
- Nenonen, Marja & Niemi, Jussi (1999) Morphological isolates in idioms: Cranberries or real words? *Brain and Language* 68: 158–164.
- Nenonen, Marja, Niemi, Jussi & Laine, Matti (2002) Representation and processing of idioms: Evidence from aphasia. *Journal of Neurolinguistics* 15: 43–58.
- Nenonen, Marja and Niemi, Sinikka (eds.) (2007) *Collocations and Idioms 1: Papers from the First Nordic Conference on Syntactic Freezes, Joensuu, May 19–20, 2006*. Studies in Languages 41, University of Joensuu. Joensuu: University of Joensuu.
- Newman, John (1996) *Give: A Cognitive Linguistic Study*. Berlin: Mouton de Gruyter.
- Niemi, Jussi (2007) Structural changes and meaning extensions of ‘pull+comp’ constructions and idioms in English and Finnish. In Marja Nenonen & Sinikka Niemi (eds.), pp. 227–234. Joensuu: University of Joensuu.
- Niemi, Jussi, Nenonen, Marja & Penttilä, Esa (1998) Number as a marker of idiomaticity. In Timo Haukioja (ed.), *Proceedings of the XVIth Scandinavian Conference of Linguistics, Turku/Åbo, November 14–16, 1996*, pp. 293–304. Turku: Åbo Akademis Tryckeri.
- Niemi, Sinikka (2004) Kielitypologisia ja kognitiivisia havaintoja ruumiinosaidiomeista. [Typological and cognitive observations on body part idioms]. In Marja Neno-

- nen (ed.), *Papers from the Finnish Conference of Linguistics, Joensuu, May 15–16, 2003*, pp.144–151. Studies in Languages 39, University of Joensuu. Joensuu: University of Joensuu.
- Niemikorpi, Antero (1991) *Suomen sanaston dynamiikkaa* [The dynamics of Finnish vocabulary]. Acta Vasaensia 26. Kielitiede 2. Vaasan yliopisto, Vaasa.
- Nunberg, Geoffrey, Sag, Ivan A. & Wasow, Thomas (1994) Idioms. *Language* 70: 491–538.
- Nykysuomen sanakirja*, Parts 1–6 (1951–1961) [Dictionary of Modern Finnish]. Helsinki, WSOY.
- O’Grady, William (1998) The syntax of idioms. *Natural Language and Linguistic Theory* 16: 279–312.
- Penttilä, Esa (2006) *It Takes an Age to do a Chomsky: Idiomaticity and Verb Phrase Constructions in English*. Unpublished doctoral dissertation, University of Joensuu.
- Reichstein, A. D. (1973) Zur Analyse der festgeprägten Sätze im Deutschen. *Deutsch als Fremdsprache* 10, 212–222.
- Rosch, Eleanor (1978) Principles of categorization. In Eleanor Rosch & Barbara B. Lloyd (eds.), *Cognition and Categorization*, pp. 27–48. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Ruhl, Charles (1999) Monosemic take. *The 25th LACUS Forum 1998*. Fullerton, CA: Linguistic Association of Canada and the United States, pp. 213–222.
- Saukkonen, Pauli, Haipus, Marjatta, Niemikorpi, Antero & Sulkala, Helena (1979) *Suomen kielen taajuussanasto* [The Finnish frequency dictionary]. Porvoo: WSOY.
- Stubbs, Michael (1986) *Educational Linguistics*. Oxford: Blackwell.
- Taverner, Theodore G. (1977) *Idioms and Idiom-Forming Verbs: A Case Grammar Analysis*. Dissertation: Georgetown University. Ann Arbor: University Microfilms International.
- Taylor, John R. (1998) Syntactic constructions as prototype categories. In Michael Tomasello (ed.), *The New Psychology of Language: Cognitive and Functional Approaches to Language Structure*, pp. 177–202. Mahwah, NJ: Laurence Erlbaum.
- Taylor, John R. (2002) *Cognitive Grammar*. New York: Oxford University Press.
- Taylor, John R. (2003) *Linguistic Categorization*. Third edition. Oxford: Oxford University Press.
- Vainio, Seppo & Nenonen, Marja (2007) Reading idioms in Finnish: An eye-tracking study. In Marja Nenonen & Sinikka Niemi (eds.), pp. 363–377. Joensuu: University of Joensuu.
- Vilkuna, Maria (1989) *Free Word Order in Finnish: Its Syntax and Discourse Functions*. Helsinki: SKS.
- Virtanen, Patrik & Pajunen, Anneli (2000) *ContextMill*. Computer software. General Linguistics, University of Turku.
- Wood, Mary McGree (1986) *A Definition of Idiom*. Bloomington: Indiana University Linguistics Club.

Contact information:

Marja Nenonen
Linguistics
University of Joensuu
P.O. Box 111
FI-80101 Joensuu, Finland
e-mail: marja(dot)nenonen(at)joensuu(dot)fi
<http://cc.joensuu.fi/linguistics/>