

Canonical Forms of Idioms in Online Dictionaries

Alenka Vrbinc

alenka.vrbinc@ef.uni-lj.si

University of Ljubljana, Slovenia

Marjeta Vrbinc

marjeta.vrbinc@ff.uni-lj.si

University of Ljubljana, Slovenia

ABSTRACT

This study focuses on the inclusion of canonical forms of idioms in different online dictionaries, and occasionally a parallel is drawn to canonical forms in print dictionaries used in the compilation of our database, which contains 141 idioms originating from literary works, ancient legends, fables and the Bible. In the foreground are dictionary users, especially learners of English, who may face a range of problems concerning canonical forms. However, we have to limit ourselves to a certain extent, which means that attention is paid mostly to the following: the use of the article preceding a noun that is the first constituent element in the idiom, the way of including information on possessives in idioms, the use of the infinitive marker, which is obligatory in some idioms beginning with a verb, the way of indicating variations in idioms, differences in the use of the apostrophe, the inclusion of similes with the comparison marker *as*, and lower- or upper-case initial letter. Analysis of the idioms from our database shows that the canonical forms may pose problems for lexicographers, who may not be sufficiently consistent, as well as to dictionary users, who may find it difficult to interpret the idiom correctly and consequently use it correctly. A possible variation should be indicated unambiguously and clearly, so as to make dictionary users aware that some idioms allow some flexibility in their form. If the comparison of the way a particular idiom is included in different dictionaries shows any differences, these are discussed and commented upon, suggestions are made and guidelines are proposed for improving the level of consistency, thus resulting in a more consistent as well as uniform and possibly more user-friendly inclusion of idioms.

Keywords: idioms; canonical forms; dictionaries of idioms; online dictionaries; learners of English

INTRODUCTION

When including lexical items in a dictionary, the first thing a lexicographer must decide on is the most typical realization of the lexical item, which is referred to as the canonical form or the dictionary citation form. In a dictionary, the lemma is included in a canonical form, and multi-word lexical items are no exception. According to Riehemann (2001, p. 32), the canonical form of the idiom is “a particular fixed phrase [...] which is recognized by speakers of the language as the normal form this idiom takes, and which is used much more frequently than would be predicted from independent factors”. In the case of multi-word lexical items, especially idioms, the determination of the canonical form is much more problematic than in single-word items, since we should be aware that quite a few idioms may appear in an altered form, since they can be subject to morphological and lexical variation, thus causing the user more serious problems (cf. also Miller, 2013; Al-Haj, Itai & Winter, 2014). This is in line with Philip (2008, p. 97), who highlights the fact that canonical forms often include grammatical elements, which inflect in text, and they may also have alternative lexical

realizations. The dictionary user may be faced with the question of how to distinguish a varied form from a canonical form, since it is the canonical form that s/he obtains in a dictionary. Besides that, it is important to mention the observation made by McGlone, Glucksberg and Cacciari (1994) that idioms in canonical forms are more readily understood than their variants. This can be explained by an approach where canonical forms have their own representation as a subtype of the idiom, which can speed up processing because they can be retrieved as pre-assembled wholes.

According to Philip (2008, p. 96), canonical forms of idioms are linguistically important because of their regularity and consistency, although corpus linguists have consistently noticed that canonical forms of idioms are not particularly common in language corpora and are outnumbered by non-canonical variants (2008, p. 95). Sinclair (1996, p. 83) agrees that fixed phrases are not in fact fixed.

Atkins and Rundell (2008, pp. 168-169) also mention different properties of idioms, such as lexical, morphological, syntactic or semantic properties, in which they stress the fact that the wording is never entirely fixed. Among the common variations, they enumerate the following: alternative words may be substituted without changing the meaning, there are parallel idioms with opposite meanings, there is no fixed canonical form, there is no complete canonical form, but there are semantic restrictions on what can fill the open slot, there are syntactic restrictions upon the idiom's behaviour, in that it undergoes only limited grammatical transformations, the idiom shows morphosyntactic flexibility, allowing inflections, agreement of possessives, etc. Fiedler (2007, p. 19-20) points out that the stable semantic and syntactic structure of a phraseological unit makes this unit distinctive from a random combination of words. This is relative stability because structural variants are possible (e.g., the difference in the use of function words or the spelling, constituent elements inside the phrase can be used in the singular or plural, lexical constituents such as nouns, verbs, adjectives, etc. can vary, grammatical and lexical elements can vary in one and the same expression). Atkins and Rundell (2008, p. 168) add that idioms are the most difficult multi-word expressions to handle in lexicography, since it is impossible to be wholly consistent because of the lack of hard and fast criteria.

The aim of this study is to focus on the inclusion of canonical forms of idioms in different online dictionaries, and occasionally a parallel is drawn to canonical forms in print dictionaries used in the compilation of our database (cf. methodology). Online dictionaries were chosen to see whether the shift from the print to electronic medium had exerted any influence on the way information is presented and accessed in a dictionary (Tan, 2008). The problems faced by a dictionary user, especially a learner of English, are identified and examined. However, in the present study attention is paid mostly to the following: the use of the article preceding a noun that is the first constituent element in the idiom, the way of including information on possessives in idioms, the use of the infinitive marker, which is obligatory in some idioms beginning with a verb, the way of indicating variations in idioms, differences in the use of the apostrophe, the inclusion of similes with the comparison marker *as*, lower- or upper-case initial letter. If the comparison of the way a particular idiom is included in different dictionaries shows any differences, these are discussed and commented upon, suggestions are made and guidelines are proposed for improving the level of consistency, thus resulting in a more consistent as well as uniform and possibly more user-friendly inclusion of idioms.

METHODOLOGY

The findings of the study presented in this contribution are based on a database containing 141 idioms that originate from literary works, ancient legends, fables and the Bible. These

idioms were compiled using two print dictionaries of idioms, i.e., *Dictionary of Idioms and Their Origins* (DITO) and *Oxford Idioms Dictionary for Learners of English* (OIDLE), which were also used as reference dictionaries. In the DITO, all the idioms that can be found in the *Index of themes* on pages 317-320 under the following titles were included: *Idioms from ancient legends*, *Idioms from the Bible*, *Idioms from fables*, *Idioms from literature* and *Idioms from Shakespeare*. The OIDLE, however, includes notes on the origin of some idioms, and only those stating clearly the literary source for the idiom were taken into consideration.

Using the Google Search “define: ...” query option, all the idioms from the database were sought in online dictionaries. Among the online dictionaries, the first three yielded by the Google Search were taken into consideration. All the information found in these three dictionaries was copied into the database and investigated to determine whether there are any differences between the canonical forms of idioms included in different dictionary sources, since it is the canonical form of idioms that is the focus of our research. The database compiled in this way makes it possible to comment on any differences that emerge from the cross-dictionary comparison. In the foreground of our research are the dictionary users, especially non-native speakers of English, since they expect to find the information they seek in the dictionary included in a way that is unambiguous and that does not raise any questions or doubts as to how this information is to be interpreted.

THE CANONICAL FORMS OF IDIOMS ACROSS THE DICTIONARY SOURCES CONSULTED

THE INCLUSION OR OMISSION OF THE ARTICLE

The canonical form of idioms whose first component element is a noun appears with either the definite or indefinite article preceding the noun in some dictionary sources in our database, whereas the article is omitted in other sources. For example:

| Example | Dictionary |
|---|--|
| <i>an albatross around one's neck</i> <i>albatross around one's neck</i> | DITO three online sources, i.e., the <i>Free Dictionary</i> , <i>Dictionary.com</i> and <i>Wiktionary</i> |
| <i>an apple of discord</i> <i>apple of discord</i> | DITO three online sources, i.e., the <i>Free Dictionary</i> , <i>Dictionary.com</i> and <i>Wiktionary</i> |
| <i>an old chestnut</i> <i>old chestnut</i> | DITO three online sources, i.e., the <i>Free Dictionary</i> , <i>Dictionary.com</i> and <i>Wiktionary</i> |
| <i>a dog in the manger</i> <i>dog in the manger</i> | DITO three online sources, i.e., the <i>Free Dictionary</i> , <i>Dictionary.com</i> and <i>Wiktionary</i> |
| <i>a/the prodigal son</i> <i>prodigal son</i> | DITO three online sources, i.e., the <i>Free Dictionary</i> , <i>Dictionary.com</i> and <i>Wiktionary</i> |
| <i>the corridors of power</i> <i>corridors of power</i> | DITO three online sources, i.e., the <i>Free Dictionary</i> , <i>Dictionary.com</i> and <i>Wiktionary</i> |
| <i>the promised land</i> <i>promised land</i> | DITO three online sources, i.e., the <i>Free Dictionary</i> , <i>Dictionary.com</i> and <i>Wiktionary</i> |

| | |
|------------------------------|--|
| <i>the salt of the earth</i> | DITO |
| <i>salt of the earth</i> | three online sources, i.e., the <i>Free Dictionary</i> , <i>Dictionary.com</i> and <i>Wiktionary</i> |
| <i>a Jekyll and Hyde</i> | DITO |
| <i>Jekyll and Hyde</i> | three online sources, i.e., the <i>Free Dictionary</i> , <i>Dictionary.com</i> and <i>Wiktionary</i> |

However, inconsistency in including idioms with or without the article can also be observed within one and the same dictionary. For example, *Kernerman English Multilingual Dictionary*, which is part of the *Free Dictionary*, lists two idioms that both begin with the noun *apple*, i.e., *apple of discord* (without the article) and *the apple of someone's eye* (with the definite article). The use of *apple of discord* is illustrated by means of an example in which *apple* is preceded by the indefinite article (*The right to host the Olympic Games is an apple of discord between the two countries*).

In some idioms, the article comes after the idiom, from which it is separated by a comma. For example, *Emperor's New Clothes*, *The* in *Longman English Dictionary Online* or *besetting sin*, *a* in DITO.

In idioms where the noun can be preceded either by the article or by the possessive pronoun, some online dictionaries omit the determiners altogether. For example, the idiom *a/your pound of flesh* is included without any determiner in three online dictionary sources (i.e., *Free Dictionary*, source: *McGraw-Hill Dictionary of American Idioms and Phrasal Verbs*, *Merriam-Webster Online* and *Phrase Finder*). Similarly, in the idiom *a/somebody's place in the sun*, the option with the possessive is indicated only in the *Free Dictionary* (source: *Cambridge Idioms Dictionary*), whereas in *Cambridge Dictionaries Online*, the same idiom appears without *somebody's* although the only example of use provided includes the possessive pronoun (i.e., *He certainly earned his place in the sun*).

POSSESSIVES IN IDIOMS

According to Moon (1998, p. 101), around 14% of the fixed expressions contained in her database (which she describes on pages 44-46) consist of slots fillable by possessive pronouns or possessive forms of nouns which cue the fixed expression deictically into context. This figure corresponds to the findings from our study, since 18 out of 141 idioms (i.e., 12.8%) included in our database also contain slots filled by possessive pronouns or possessive forms of nouns. The possessives in the idioms are represented by either *one's* (when the possessive refers to the same person as the subject) or *sb's* (when the people referred to by the possessive pronoun are different from the subject). Atkins and Rundell (2008: 362–363) also address the issue of idioms containing possessives and state that the correct canonical forms are those including *one's* and *someone's*. In recent editions of English monolingual learner's dictionaries, the generic *one's* is mostly replaced by *your*, also implying the entire range of possessive pronouns, for example:

| Idiom | Dictionary |
|---|--|
| <i>an albatross around one's neck</i> | DITO, <i>Dictionary.com</i> and <i>Wiktionary</i> |
| <i>albatross around/round your neck</i> | <i>Free Dictionary</i> |
| <i>the scales fall from somebody's eyes</i> | OIDLE, <i>Free Dictionary</i> and <i>Cambridge Dictionaries Online</i> |

As a matter of fact, *your* is now frequently used as the generic possessive, which may illustrate the assumption or presupposition on the part of the speaker that his/her interlocutor possesses certain things (Fellbaum, 1993: 282). Nevertheless, it is also possible to come across idioms with other possessive pronouns (e.g., *the apple of my eye* in *Phrase Finder*). However, several dictionaries do not follow these rules consistently, for example:

| | |
|---|--------------------------------------|
| Idiom | Dictionary |
| <i>there's method in sb's madness</i> | OIDLE and <i>Free Dictionary</i> |
| <i>there is method in one's madness</i> | DITO |
| <i>have method in your madness</i> | <i>Cambridge Dictionaries Online</i> |
| <i>there's method in my madness</i> | <i>Phrase Finder</i> |

THE INFINITIVE MARKER *TO* IN IDIOMS

As has been established, idioms are usually presented in the base form, e.g., an infinitive phrase if the idiom is a verb phrase. This is also in line with the observations made on the basis of those idioms included in our database, but one point should be emphasized: that there is some discrepancy in the treatment of idioms beginning with verbs. In the DITO, the verbs are preceded by the infinitive marker *to* as opposed to all other dictionary sources found in our database, where *to* is not used. For example:

| | |
|--------------------------------|--|
| Idiom | Dictionary |
| <i>to have an axe to grind</i> | DITO |
| <i>have an axe to grind</i> | OIDLE, <i>Free Dictionary</i> , <i>Phrase Finder</i> , <i>Cambridge Dictionaries Online</i> |
| <i>not to have a clue</i> | DITO |
| <i>not have a clue</i> | OIDLE, <i>Free Dictionary</i> , <i>Cambridge Dictionaries Online</i> , <i>YourDictionary</i> |

As far as the idiom (*to*) *add insult to injury* is concerned, it should be stressed that the infinitive marker *to* is used not only in the DITO but exceptionally also in *Cambridge Dictionaries Online*, in which the infinitive marker *to* is not normally used to precede verbs at the beginning of idioms. At first sight, this idiom seems to be an exception, but an explanation can easily be found if we take a closer look at the definition (i.e., 'said when you feel that someone has made a bad situation worse by doing something else to upset you'). The purpose of this definition is to explain the pragmatic message carried by the idiom. In addition, the example illustrating its use typically shows this pragmatic use (i.e., *They told me I was too old for the job, and then to add insult to injury, they refused to pay my expenses!*). Interestingly, some other dictionary sources from our database include similar examples of use with the same form of the idiom as in *Cambridge Dictionaries Online*, for example:

| | |
|--|---|
| Dictionary | Example |
| <i>Free Dictionary</i> , source <i>McGraw-Hill Dictionary of American Idioms and Phrasal Verbs</i> | <i>First, the basement flooded, and then, to add insult to injury, a pipe burst in the kitchen. My car barely started this morning, and to add insult to injury, I got a flat tire in the driveway.</i> |
| <i>Free Dictionary</i> , source <i>Cambridge Idioms Dictionary</i> | <i>First of all he arrived an hour late and then, to add insult to injury, he proceeded to complain about my choice of restaurant.</i> |

Wiktionary

As if the hostile takeover weren't enough, to add insult to injury they scrapped ninety percent of our products and replaced them with their own.

These dictionary sources, however, do not use the infinitive marker *to* and they use the infinitival definitions, thus not referring to the pragmatic use, for example:

Dictionary

Free Dictionary, source *McGraw-Hill Dictionary of American Idioms and Phrasal Verbs*

Free Dictionary, source *Cambridge Idioms Dictionary*

Wiktionary

Definition

to make a bad situation worse,
to hurt the feelings of a person who has already been hurt

to make a bad situation even worse for someone by doing something else to upset them

to further a loss with mockery or indignity,
to worsen an unfavourable situation

THE INCLUSION OF VARIATIONS IN IDIOMS

Fixedness is definitely a vital feature of idioms, and many idioms do not vary at all, but as Moon (1998, pp. 120-121) observes when studying the fixed expressions and idioms included in her database, around 40% of these items have lexical variations or strongly institutionalized transformations, and around 14% have two or more variations on their canonical forms. The figure referring to lexical variations or strongly institutionalized transformations corresponds to the findings from our study, since 44 out of 141 idioms (i.e., 31.2%) included in our database also contain lexical variations. Alternative words may be substituted without changing the meaning of the idiom (Atkins & Rundell, 2008, p. 168; Svensén, 2009, p. 189), and as Moon (1998, p. 122) points out, “the variant forms of individual expressions are to be considered as variations rather than as separate expressions with coincidentally the same meaning and with some lexis in common”. The fixedness of idioms should not necessarily be taken for granted and variant forms do exist, although “variants are embellishments, additions or reductions whose meaning and function are ultimately dependent on, and reducible to, the canonical form from which they are derived”, as Philip (2008, p. 96) puts it. In dictionaries, variations are mostly shown by using either a slash or a comma. It should, however, be stressed that compilers of a dictionary have to be very careful how to include variations, since the comma is not normally used to indicate variation and should therefore be avoided in this function, whereas the slash is used to separate alternative words, which means that the word following the slash can replace only the word preceding it. For example:

Idiom

cover/hide a multitude of sins
raise Cain/hell

Dictionary

OIDLE, *Macmillan Dictionary Online*
OIDLE

However, a misleading use of the slash or the comma used to indicate variation can be found in various dictionaries. For example:

| Idiom | Dictionary |
|---|--|
| <i>castle in the air/in Spain</i> | DITO |
| <i>castle in the air, in Spain</i> | <i>Free Dictionary</i> |
| <i>the patience of a saint/of Job</i> | OIDLE, <i>Macmillan Dictionary Online, Free Dictionary</i> |
| <i>put on, wear, etc. sackcloth and ashes</i> | OIDLE, <i>OALD9 Online</i> |
| <i>keep to, stay on, etc. the straight and narrow</i> | OIDLE |

NOUN VARIATIONS WITH SINGULAR OR PLURAL FORMS

In noun variations, the noun may take either the singular or the plural form. In studying the idioms included in our database, it can be established that in some idioms this type of variation is not indicated, since in some dictionary sources only the singular form is used, whereas in others, the same idiom appears only with the plural noun. For example:

| Idiom | Dictionary |
|-----------------------------------|---------------------------------|
| <i>castle in the air/in Spain</i> | DITO and <i>Free Dictionary</i> |
| <i>(build) castles in the air</i> | OIDLE |
| <i>build castles in the air</i> | <i>Wiktionary</i> |
| <i>castles in the air</i> | <i>Dictionary.com</i> |

The idiom *to kill the goose which lays the golden egg/eggs* is even more variously represented in the dictionaries in our database, e.g.:

| Idiom | Dictionary |
|---|---|
| <i>(to kill) the goose which lays the golden eggs</i> | DITO |
| <i>kill the goose that lays the golden egg</i> | <i>Free Dictionary, source: Cambridge Idioms Dictionary</i> |

However, it should be stressed that in other sources, the noun variation is clearly shown either by means of a slash or by providing the plural -s in brackets. For example:

| Idiom | Dictionary |
|---|---|
| <i>(kill) the goose that lays the golden egg/eggs</i> | OIDLE |
| <i>kill the goose that lays the golden egg(s)</i> | <i>Free Dictionary, source: McGraw-Hill Dictionary of American Idioms and Phrasal Verbs</i> |

DIFFERENCES IN THE USE OF THE GENITIVE

As far as the idiom *the lion's den* (OIDLE, *Cambridge Dictionaries Online* and *Wiktionary*) vs *the lions' den* (*Free Dictionary, source: Cambridge Idioms Dictionary*) is concerned, comparison of the different dictionary sources included in our database reveals that the singular form is prevalent, although the plural form, i.e., *the lions' den*, may seem to be more logical if etymology is taken into account. It is known that Daniel was thrown into a pit of lions (cf. Contemporary English Version of the Bible, Daniel 6,1–28) or into a den of lions (cf. King James Version, Daniel 6,1–28). If the frequency of either form of this idiom is

checked in the BNC and ukWac, it should be pointed out that the number of hits for *lion's den* is much higher than that for *lions' den*.

| | BNC | ukWac |
|-------------------|-----|-------|
| <i>lion's den</i> | 15 | 107 |
| <i>lions' den</i> | 6 | 53 |

In the three online dictionaries (*Phrase Finder*, *Free Dictionary*, *Merriam-Webster Online*) as well as in both reference dictionaries (DITO, OIDL), the idiom (*an/sb's*) *Achilles' heel* is spelt with an apostrophe, which means that the dictionaries agree as to the spelling of this idiom. This spelling is expected and should be considered the only acceptable, since it is in accordance with grammar rules according to which the zero genitive is used with Greek names of more than one syllable which end in *-s*. Interestingly, no hits for *Achilles' heel* can be found in the BNC, whereas in the ukWac there are 11 hits. In both corpora, the spelling without the apostrophe (*Achilles heel*) yields more hits (BNC 33 hits, ukWac 444 hits).

THE INCLUSION OF SIMILES WITH THE COMPARISON MARKER *AS*

Similes are structurally fixed and draw upon comparisons between two very different but explicit entities. Two frequent types can be distinguished, i.e., (*as*) + adjective + *as* + noun phrase and (verb) + *like* + noun phrase (Fiedler, 2007, pp. 43-44; cf. also Scotto di Carlo, 2014), and both of these can be found in our database. Here, we will focus only on the first type, since the comparison marker *as* is an optional component and is treated differently in different dictionary sources: in some of them, it is included as if it were the obligatory component, i.e., *as rich as Croesus* (DITO and *Phrase Finder*), in others it is made clear that it is optional to use *as*, i.e., (*as*) *rich as Croesus* (OIDL), whereas in some dictionary sources there is no indication that *as* may be a constituent element of the simile, i.e., *rich as Croesus* (*Dictionary.com* and *YourDictionary*). A similar treatment can be observed in the simile *as mad as a hatter* (DITO, *Phrase Finder* and *Free Dictionary*, in which the simile is preceded by the verb *be*), (*as*) *mad as a hatter* (OIDL) and *mad as a hatter* (*Urban Dictionary*).

LOWER- OR UPPER-CASE INITIAL LETTER

Another observation made when carrying out an in-depth study into the form of idioms compiled in our database concerns the use of a lower- or upper-case initial letter in some idioms. As can be seen, there are some idioms in which the use of a lower-case or upper-case initial may vary. For example:

| Idiom | Dictionary |
|---------------------------------|--|
| <i>Man/Girl Friday</i> | DITO |
| <i>girl Friday</i> | <i>YourDictionary</i> |
| <i>a girl/man/person Friday</i> | <i>Free Dictionary</i> |
| <i>Parkinson's law</i> | OIDL, <i>Free Dictionary</i> , source: <i>Collins English Dictionary</i> |
| <i>Parkinson's Law</i> | <i>Free Dictionary</i> , source: <i>The American Heritage Dictionary of the English Language</i> , <i>Merriam-Webster Online</i> |
| <i>Parkinson's law (or Law)</i> | <i>Free Dictionary</i> , source: <i>Random House Kernerman Webster's College Dictionary</i> |

(the) promised land
Promised Land

OIDLE, Merriam-Webster Online
Free Dictionary, source: *Collins English Dictionary, Collins Dictionaries*

Promised Land is presented as the headword in the *Free Dictionary* (source: *The American Heritage Dictionary of the English Language*) and is capitalized, i.e., *Promised Land*, but the dictionary article is polysemous in that sense 1 explains the original non-metaphorical meaning, while sense 2 refers to the metaphorical meaning but is not capitalized.

In all examples listed above, the spelling of at least one idiom component varies. The capitalization of the component suggests a proper name. If, on the other hand, the component is spelt with a lower-case initial, the only interpretation possible is that this component should be comprehended as a common noun rather than a proper noun.

DISCUSSION

Studying the canonical forms of idioms in dictionaries reveals that different dictionaries or even one and the same dictionary may not consistently employ the same principles for the inclusion of idioms in their canonical form. No criticism is justifiable if it can be established that one dictionary uses one policy consistently, but on closer observation, it is obvious that some (though not all) dictionaries do not stick to the same policy (cf. *Cambridge Idioms Dictionary: an accident waiting to happen, an ace in the hole vs albatross around/round your neck; DITO: a fly on the wall vs forlorn hope, a*). An inter-dictionary comparison reveals that different dictionaries use different policies, which probably does not pose problems for native speakers of English, who are well acquainted with the syntactic characteristics of their mother tongue, but may be less appropriate for non-native speakers, who would benefit from obtaining clear guidance on the form and use of an idiom or other lexical items. One of the main problems concerning the online dictionaries is the absence of information about the target users of the dictionary in question, which means that anybody with internet access can consult an online dictionary, be it a native or a non-native speaker. Although online dictionaries may be used by native speakers of English without any particular problems, non-native speakers may experience difficulty. In traditional lexicography, a clear demarcation line exists between dictionaries intended for native speakers and those intended for learners, and the same principle should also be applied to online dictionaries.

As shown in the canonical forms of idioms across the dictionary sources consulted, various inconsistencies in the canonical forms of idioms can be found; these issues are discussed below and recommendations as to possible improvements are formulated.

INCLUSION OF ARTICLES

The first issue presented in the section on the canonical forms of idioms across the dictionary sources consulted is the inconsistent inclusion of the article, which can be observed in online dictionaries as well as in both reference dictionaries. Native speakers of English would probably not even consider this a problem, but non-native speakers will find it difficult to decide whether to use an article or not, especially if articles are not used in their mother tongue (e.g., in Slavonic languages). In all the cases mentioned in the inclusion or omission of the article, it would be advisable to include the article as part of the canonical form. Apart from that, it proves valuable if examples of use are provided, so that non-native speakers can check how the article is used, for example:

Example

That old car is an albatross around my neck

The failed real estate scheme became an albatross around her neck, for now she could not interest other investors in a new project

Dictionary

American Heritage New Dictionary of Cultural Literacy (Dictionary.com)

American Heritage Dictionary of Idioms (Dictionary.com)

In idioms containing proper nouns (e.g., *a Jekyll and Hyde*), users who are not linguists should be made aware of the fact that the proper name has undergone the process of appellativization and is now used as a common noun; thus, the use of the article is obligatory, which should be clearly indicated.

As far as the inclusion of articles is concerned, some dictionaries omit the article in initial position and place it at the end of the idiom after a comma (e.g., *Emperor's New Clothes, The* or *besetting sin, a*). A user is supposed to interpret the correct order of the component elements of an idiom. This practice is to be avoided because it is much clearer for users when the article appears in the place where it is actually used in speaking or writing. It is not just the article that appears at the end of the idiom from which it is separated by a comma, but one idiom has been found in our database where the verb comes after the comma, i.e., *full circle, come (YourDictionary)*. If read as such, it makes no sense, but the question can be raised whether an average dictionary user, especially a non-native speaker, would be able to decipher the correct word order of the idiom and consequently use the idiom correctly in context.

POSSESSIVE PRONOUNS

In idioms where there is an option to choose between the article and the possessive pronoun preceding a noun (e.g., *a/your pound of flesh*), it can also not be considered a good policy to omit the article or the possessive pronoun altogether. This piece of information is certainly needed at least by non-native speakers using the dictionary and should therefore be provided.

All in all, it would be advisable to indicate whether the noun at the beginning of the idiom should be preceded by the article, to make a clear distinction between such idioms and those where the article is not used. Such a method for including the idioms would be useful to non-native speakers, who may find it difficult to use the appropriate article with a particular noun, especially if we bear in mind that the use of the article in English is often idiomatic and therefore highly unpredictable for non-native speakers.

The next issue that should be addressed is the indication of possessives. Frequently, the generic *one's* or *your* is used to indicate that the user should select the appropriate possessive pronoun that fits into his/her context. This means that the dictionary user whose mother tongue is not English is faced with a difficult task, especially if the possessive pronoun in English is not used in the way typical of his/her mother tongue. It should be stressed that, if we compare English possessive pronouns and, for example, Slovene possessive pronouns, we can observe that the English pronouns are characterized by their incomplete conformity with the Slovene possessive pronoun *svoj*, which does not change its form contextually; in other words, it remains unchanged in its basic form (but not as regards declension) regardless of the subject of the sentence. German, on the other hand, shows a complete conformity with English, since the generic possessive pronoun used in the canonical forms is the third person singular masculine *sein*, which means that in context, the generic *sein* is replaced by the entire set of possessive pronouns depending on the subject of the sentence.

In the editions of British monolingual learner's dictionaries published before the 1990s, the possessive pronoun is presented as *one's*. Since that time, it has been replaced by the possessive pronoun *your*, which can be misleading to many target users of these dictionaries, i.e., non-native speakers of English. Given that detailed front matter has mostly been abandoned in dictionaries, users are not even made aware of the fact that the possessive pronoun *your* in idioms may pose problems, since in the majority of cases *your* should be replaced by an appropriate possessive pronoun. The only hints they may get can be found in specialized print idioms dictionaries, such as the following:

- *Collins Cobuild Dictionary of Idioms* (2002: x): “[W]e use words such as ‘you’, ‘your’, ‘yourself’, and ‘their’ to indicate that an appropriate pronoun or possessive adjective should be supplied.” or
- *Cambridge Idioms Dictionary* (2006: ix): “‘your’ can be replaced by ‘his/her/their/our/my’.”

However, it is questionable whether users read the front matter and interpret the explanation correctly. A more comprehensive explanation is provided in the *American Heritage Dictionary of Idioms*, 2nd edition. In *Guide to Using This Book*, a section entitled *Variable Pronouns* is dedicated to this issue, and users learn when or how *one*, *someone*, *one's* and *someone's* are used in idioms. Their use is explained as follows:

“Many idioms can be used with different pronouns [...]. Consequently, the pronouns *one* and *someone* are used in the entry words and variants to indicate that the object or possessive pronoun in the idiom may vary according to context. *One* or *one's* means that the antecedent of the pronoun must be the subject of the clause, or in some cases an inanimate noun or a gerund must be the subject. [...]

The use of *someone* or *someone's* in the idiom means that the pronoun can be replaced only by a noun or pronoun that does not refer to the grammatical subject of the clause. In other words, the action of the verb is directed from one person to another (the “someone”). [...]

In some cases, however, *your* is a fixed element of the idiom and should be retained in its unchanged form (e.g., *as long as your arm*; *thank your lucky stars*; *a penny for your thoughts*; *use your head/loaf*; *zip your lip*; *Bob's your uncle*; *wash your mouth out*; *not on your life* – these examples are not taken from the database compiled for the purpose of this study). The question can be raised how users can know when *your* in an idiom is replaceable and when it represents a compulsory constituent element. In order to avoid this dilemma, it would be advisable to include the English idioms in (mono- as well as bilingual) dictionaries in two ways:

- 1) with *one's* where the possessive pronoun should be replaced by the appropriate possessive pronoun, since the possessive pronoun must agree with the subject of the verb or with *somebody's* (and not *your*) if it does not agree with the subject;
- 2) with *your* if *your* is obligatory and cannot be replaced by any other possessive pronoun.

THE USE OF PARTICLE *TO*

Another issue that captured our attention is the use of the particle denoting the infinitive. If it forms part of the canonical form of the idiom, a dictionary user may rightly interpret it as being part of the idiom and thus an element that should not be omitted. The real problem here is that the infinitive marker *to* can be an obligatory element in some idioms – in idioms that are used in the infinitival form only. In our database, no idioms with the infinitive marker *to*

as an obligatory element can be found, but when going through the idioms in the OI_{DL}E, in which idioms beginning with verbs do not contain *to*, the only exception being idioms where *to* is obligatory, the following nineteen idioms beginning with the infinitive marker *to* can be found: *to begin with*, *to coin a phrase*, *to come*, *to die for*, *not to put too fine a point on it*, *to cut/make a long story short*, *to look at sb/sth*, *not to mention sth*, *to be (more) precise*, *to be on the safe side*, *not to say*, *to say the (very) least*, *to say nothing of sth*, *not to be sneezed/sniffed at*, *to spare*, *to be sure*, *to tell (you) the truth*, *to think (that)*, *not to worry*.

Taking into account that some idioms are used only in the infinitival form, while others appear in the infinitive only to fulfil the criteria for the canonical form, it is advisable to adopt the same policy as the OI_{DL}E and to use the infinitive marker *to* at the beginning of the idiom only if it is an obligatory element.

INCLUSION OF VARIATION

A very important feature regarding the inclusion of phraseological units in dictionaries is how to include variation in the way that is least ambiguous and most user-friendly. The slash is the most common symbol used to show two or more alternatives and if used properly, it is the easiest way of marking variation. If, on the other hand, the slash is used incorrectly (as is frequently the case in different dictionaries), users may have problems determining the entire form of an idiom because it can be difficult for them to decipher which element can fill a certain slot. This is why the slash should be used to indicate the variability in such a way that one word to the left of the slash and one word to the right of the slash can fill one slot. If, for example, a dictionary user comes across the idiom *castle in the air/in Spain*, s/he could easily interpret the entire idiom as **castle in the air Spain* or **castle in the in Spain*, both of which are incorrect. In such cases, it would be recommendable to avoid the use of the slash and use ‘or’ instead: *castle in the air* or *castle in Spain*.

In cases where variation is more open, dictionaries typically use ‘etc.’ to indicate that the options enumerated can be further expanded or, in other words, that there are obviously more alternatives than suggested (e.g., *put on*, *wear*, etc. *sackcloth and ashes* or *keep to*, *stay on*, etc. *the straight and narrow*). At least from the point of view of those dictionary users who are not native speakers, the practice of using ‘etc.’ should be avoided, because its function is questionable, since it does not tell the users anything about what other alternatives are possible.

In idioms containing nouns that can be used in either the singular or plural form, the inclusion of both noun variants should be recommended (e.g., *(kill) the goose that lays the golden egg/eggs* or *kill the goose that lays the golden egg(s)*), otherwise a dictionary user may justifiably presuppose that only the singular or only the plural form is possible (as is also the case in a great number of idioms). However, s/he should be made aware of the fact that both forms of the noun can indeed be used in context.

GENITIVE INFLECTION

As regards the use of the genitive inflection, the form of nouns differs in different dictionaries. The dictionaries consulted for the purpose of this study do not offer both the singular and the plural forms but limit themselves to giving one form only (e.g., *the lion’s den* or *the lions’ den*). In such cases, one would rightly expect the dictionary to list the form with a higher frequency of occurrence and a corpus search confirms this presupposition (as is evident from the results in the section Differences in the use of the genitive). As far as the idiom *Achilles’ heel* is concerned, one can only wonder why there is such discrepancy between the dictionaries (print as well as online), on the one hand, and the corpora, on the

other, given that all dictionaries nowadays claim to be based on data from various corpora. If this were in fact true, the idiom should be included in dictionaries without the apostrophe.

SIMILES

As for the inclusion of similes, the confusion regarding the use or the omission of the comparison marker *as* is totally unnecessary. The fact that the first *as* in a simile is optional can best be indicated by putting *as* in brackets. In this way, the dictionary user, especially a non-native speaker, would clearly be made aware of the fact that s/he can either use or omit *as* when using the simile in context.

INITIAL LETTER

The issue of using a lower- or upper-case initial letter addressed in the section Lower- or upper-case initial letter demands closer attention. A name can change from a proper name to an appellative when a proper name takes on a metaphorical meaning. As Svensén (2009, p. 74) points out, the metaphorical nature is often lost, and the connotations once associated with a certain name are known only to those familiar with its original meaning. He also claims that appellativization can be more or less ‘forgotten’, which may be manifested by the common noun being written with a lower-case initial letter. This is why originally proper nouns can be spelt with the lower-case initial letter, but the examples shown in the section Lower- or upper-case initial letter are suggestive of some other problems. The idiom *man/girl Friday* is variously included in dictionaries in that some dictionaries include it by capitalizing the first noun in the idiom (*Man/Girl*). Here, there is no reason whatsoever to spell *man/girl* with the capital initial, since these two nouns are common rather than proper nouns, which can be seen from the definition for *man Friday* in the OED: ‘a servile follower or attendant; a factotum or servant of all work. (The allusion is to Robinson Crusoe’s servant, whom he usually refers to as ‘my man Friday.’ The same spelling as in *man Friday* should also be applied to *girl Friday*, since it was coined to describe ‘a female assistant’ (DITO, p. 194).

The idiom *Parkinson’s law/Law* is more complicated, since the capitalization of the noun *law/Law* depends on the interpretation of the entire idiom. If *Parkinson’s Law* is to be interpreted as a proper noun, the use of the capital initial in *law* is justifiable. If, on the other hand, *Parkinson’s law* is not considered to be the name of the law, which is a plausible interpretation given that the lexical item is an idiom proper and as such used in the idiomatic meaning, *law* is logically spelt with a lower-case initial. If one idiom can be used metaphorically and non-metaphorically (e.g., *Promised Land*), the spelling depends on the meaning. Usually, the original, i.e., the non-metaphorical meaning, is capitalized, while the metaphorical one has acquired the status of a common noun. It is therefore spelt with the lower-case initial. If this is the case, a dictionary should make the difference absolutely clear by indicating the correct spelling for each individual sense.

CONCLUSION

As indicated based on the analysis of the idioms included in our database, the canonical forms are far from being clear-cut. It seems that they pose problems for both lexicographers and users. Lexicographers are not consistent in the inclusion of canonical forms, and users are faced with challenges concerning the interpretation and consequently the correct use of the idiom in question. Doubtlessly, some idioms are more fixed as regards their form than others, and some show a certain degree of variation, although it should be stressed that the type of variation a given idiom can undergo may be unpredictable, especially for a non-native speaker. As has already been stressed, it is essential to indicate any variation as

unambiguously and clearly as possible, so as to make potential dictionary users aware that some idioms allow some flexibility in their form. It can be presupposed that idioms are looked up more often by non-native speakers who encounter them in a certain context and want to learn what they mean, or who want to check the correct form of a certain idiom they intend to use when writing or speaking. Therefore, great care should be taken by lexicographers to ensure that the idiom is included in such a way that the user gets complete information.

To conclude, it can be established with a high degree of certainty that there is not much difference between the inclusion of idioms in print and online dictionaries, and the same deficiencies, however slight they may seem, can be found in either type of dictionary. There is still room for improvement; therefore, some guidelines have been proposed in this contribution that can be applied to make the inclusion of idioms more consistent and more user-friendly. The same guidelines can also be followed in general monolingual learner's dictionaries and especially in bilingual dictionaries where the intended user is a foreign learner who needs guidance on different aspects and the canonical form of idioms is no exception.

REFERENCES

- Al-Haj, H., Itai, A. & Winter, S. (2014). Lexical Representation of Multiword Expressions in Morphologically-complex Languages. *International Journal of Lexicography*. 27(2), 130-170.
- Ammer, C. (2013). *The American Heritage Dictionary of Idioms*. Second Edition. Boston: Houghton Mifflin Harcourt.
- Atkins, B.T.S. & Rundell, M. (2008). *The Oxford Guide to Practical Lexicography*. Oxford: Oxford University Press.
- Cambridge Dictionaries Online*. Retrieved June 20, 2015 from <http://dictionary.cambridge.org/>
- Dictionary.com*. Retrieved June 20, 2015 from <http://dictionary.reference.com/>
- Fellbaum, C. (1993). The Determiner in English Idioms. In C. Cacciari & P. Tabossi, (Eds.). *Idioms: Processing, Structure, and Interpretation* (pp. 271-295). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Fiedler, S. (2007). *English Phraseology*. Tübingen: Gunter Narr Verlag.
- Flavell, L. & Flavell, R. (2006). *Dictionary of Idioms and Their Origins*. London: Kyle Cathie Limited. (DITO)
- Free Dictionary*. Retrieved June 20, 2015 from <http://www.thefreedictionary.com/>
- Kim Hua, T. & Woods, P. C. (2008). Media-Related or Generic-Related Features in Electronic Dictionaries: Learner's Perception and Preferences. *GEMA Online® Journal of Language Studies*. 8(2), 1-17.
- Longman English Dictionary Online*. Retrieved June 20, 2015 from <http://www.ldoceonline.com/>
- Macmillan Dictionary Online*. Retrieved June 20, 2015 from <http://www.macmillandictionary.com/>
- McGlone, M.S., Glucksberg, S. & Cacciari, C. (1994). Semantic Productivity and Idiom Comprehension. *Discourse Processes*. 17, 167-190.
- Merriam-Webster Online*. Retrieved June 20, 2015 from <http://www.merriam-webster.com/>
- Miller, J. (2013). Birds of a Feather Don't Always Flock Together: User Problems in Identifying Headwords in Online English Learner's Dictionaries. *Lexikos*. 23, 273-285.

- Moon, R. (1998). *Fixed Expressions and Idioms in English: A Corpus-Based Approach*. Oxford: Clarendon Press.
- OALD9 Online*. Retrieved June 20, 2015 from <http://www.oxfordlearnersdictionaries.com/definition/english/>
- Parkinson, D. (ed.). (2006). *Oxford Idioms Dictionary for Learners of English*. Second edition. Oxford: Oxford University Press. (OIDLE)
- Philip, G. (2008). Reassessing the Canon: 'Fixed' Phrases in General Reference Corpora. In S. Granger & F. Meunier (Eds.). *Phraseology: An Interdisciplinary Perspective* (pp. 95-108). Amsterdam, Philadelphia: John Benjamins Publishing Company.
- Phrase Finder*. Retrieved June 20, 2015 from <http://www.phrases.org.uk/>
- Riehemann, S.Z. (2001). A Constructional Approach to Idioms and Word Formation. Phd Thesis, Stanford University. Retrieved June 20, 2015 from <http://doors.stanford.edu/~sr/sr-diss.pdf>
- Scotto di Carlo, G. (2014). Figurative Language in Science Popularisation: Similes as an Explanatory Strategy in TED Talks. *3L: Language Linguistics Literature®*, *Southeast Asian Journal of English Language Studies*. 20(3), 1-16.
- Sinclair, J.M. (1996). The Search for Units of Meaning. *TEXTUS*. 9(1), 75-106.
- Sinclair, J., Sinclair Knight, L. & Clari, M. (eds.). (2002). *Collins COBUILD Dictionary of Idioms*. Second edition. London: HarperCollins.
- Svensén, B. (2009). *A Handbook of Lexicography: The Theory and Practice of Dictionary-Making*. Cambridge: Cambridge University Press.
- Urban Dictionary*. Retrieved June 20, 2015 from <http://www.urbandictionary.com/>
- Walter, E. (ed.). (2006). *Cambridge Idioms Dictionary*. Cambridge: Cambridge University Press.
- Wiktionary*. Retrieved June 20, 2015 from https://en.wiktionary.org/wiki/Wiktionary:Main_Page
- YourDictionary*. Retrieved June 20, 2015 from <http://www.yourdictionary.com/>

Corpora

Retrieved June 20, 2015 from <https://ca.sketchengine.co.uk/>

Other online sources

Retrieved June 20, 2015 from <http://www.biblija.net>

ABOUT THE AUTHORS

Alenka Vrbinc is Associate Professor at the Faculty of Economics, University of Ljubljana, Slovenia. She has been active in the field of lexicography, primarily learner lexicography, and also in the field of phraseology. She has edited a German-Slovene dictionary and co-authored an English-Slovene dictionary.

Marjeta Vrbinc is Associate Professor at the Faculty of Arts, University of Ljubljana, Slovenia. Her main research interests are contrastive lexicology and phraseology, practical lexicography, with the emphasis on English monolingual learner's dictionaries, and learners' dictionary skills. She has edited a German-Slovene dictionary and co-authored an English-Slovene dictionary.