Available online at www.jlls.org

JOURNAL OF LANGUAGE AND LINGUISTIC STUDIES

ISSN: 1305-578X

Journal of Language and Linguistic Studies, 18(1), 1021-1032; 2022

A Cognitive corpus-based study of exocentric compounds in English

Hicham Lahlou [a 1], Abdullah Imran Ho [b]

[a] *Universiti Sains Malaysia, Malaysia*
[b] *Universiti Kebangsaan Malaysia, Malaysia*

APA Citation:

Lahlou, H., & Ho, A.I. (2022). A Cognitive corpus-based study of exocentric compounds in English. *Journal of Language and Linguistic Studies, 18*(1), 1021-1032. Doi: 10.52462/jlls.236

Submission Date:31/08/2021

Acceptance Date:20/11/2021

Abstract

Exocentric compounding is a creative morphological process that contributes to the English lexicon. However, because it lacks a syntactic or semantic head, it was deemed an exceptional case in most word-formation literature and hence neglected. Previous work has only been limited to syntax-based grammar and the notion of headedness and thus failed to address the other linguistic rules that constrain exocentric compounds. The current paper aims to identify the frequency of exocentric compounds and thus to determine their viability. The research will also look into how conceptual metaphor and/or conceptual metonymy motivate exocentric compound formation. The results demonstrate that exocentric compounds are viable lexical units, generating content words (e.g., adjectives and adverbs) through productive word-formation processes, and extending word senses. The results also suggest that conceptual metonymy is more active than conceptual metaphor in the formation of exocentric compounds. The present findings have several implications for research on exocentric compounds, conceptual metonymy and conceptual metaphor.

Keywords: English; exocentric compounds; conceptual metonymy; conceptual metaphor; source domains; target domains

1. Introduction

The process of word formation is one of the most productive ways of expanding the English vocabulary. Language users utilise several morphological devices to generate new words, including (among others) derivation and compounding. To describe newly encountered situations, it allows a language user to construct new words from existing ones (Hamawand, 2011). Compounding, for example, has contributed significantly to modern English vocabulary (Szymanek, 2005), and thus it has attracted much attention from linguists.

Most research on word-formation has been devoted to compounds that abide by standard grammatical (or syntactic) and morphological patterns. Very little effort was made to analyse those classically deemed to be exceptional. Bloomfield's (1933) classification of compounds into distinct types, endocentric and exocentric compounds, has tremendously affected the distinction between these compound formations. An endocentric compound is one that is the hyponym of the head element, as in *homework*, a kind of work. An exocentric (headless) compound is one that is not a hyponym of the head element. Most of these compounds are generated metaphorically or metonymically, as in *land*

[1] Corresponding author.
E-mail address: hlahlou2003@hotmail.com

fishing, to use Benczes's (2006) example. The compound l*and fishing* is not a sort of fishing and has nothing to do with water or fish. It is an activity that involves using a metal detector to seek treasure hidden in the ground.

Endocentric and exocentric compounds are defined not just by semantic headedness, but also by syntactic headedness. Endocentric compounds, as opposed to exocentric compounds, obey the Righthand Head Rule (RHR), which states that the head is the right-hand component of a morphologically complex lexical unit (Williams, 1981). The head thus shows the key sense of the compound and falls under the same grammatical category as the whole compound. Consider an exocentric compound noun that comprises a verb and an adverb or preposition, i.e., with no head noun, such as *write-off*, *takeover*, *breakdown*, and *sell-out*, to use Carstairs-McCarthy's (2002) examples. Unlike endocentric compounds, exocentric compounds lack the head both semantically and syntactically. Given the concept of headedness, the literature on linguistics in general and word-formation in particular marginalised exocentric compounds.

As they were deemed to be exceptional cases not following regular and productive word-formation processes, they were hardly addressed (if ever) in the literature (Benczes, 2006). Many previous studies have stated that exocentric compounds are rare in any language (e.g., Bloomfield, 1933; Spencer, 1991; Adams, 2014). According to some studies, the meanings of exocentric compounds are not subject to compositionality and are as opaque as idioms. As a result, they should merely be listed in the lexicon (e.g., Katamba, 2006) or covered by a list rather than a set of rules (Allen, 1979). According to Katamba (2006), the opacity of exocentric compounds is the fundamental reason they are used far less frequently in the formation of new words than endocentric compounding. The absence of a head in exocentric compounds, as well as the unpredictability of their meaning, makes their nature problematic. As a result, they are sometimes considered as irregular and unproductive (e.g., Allen, 1979; Ten Hacken, 2000).

The traditional methodologies used in studies on word formation are one of the key reasons exocentric compounds were classified in the way stated above. An entity must meet all the necessary and sufficient conditions to be a member of a category under this approach, or it will be a non-member. For example, the either-or methodology simplifies the distinction between the meanings of words in componential analysis by selecting only one attribute, either - or +. Only clear-cut examples can be handled using this method. The question that arises here is what about creative contributory word-formation processes that do not meet all the conditions and violate one or more grammatical rules? Exocentric compounds, such as idioms and linguistic metaphors, enrich vocabulary and require extensive research. Being less frequently used should never be used as an excuse to ignore different linguistic phenomena, particularly when they result from cognitive motivation rather than concrete coining of morphemes. The present paper aims to determine the frequency of the selected exocentric compounds. The study examines a data set randomly chosen from the literature on exocentric compounds to broaden current knowledge of their viability. The study also aims to explore the extent to which the cognitive mechanisms, conceptual metaphor and/or conceptual metonymy motivate exocentric compound formation.

2. Literature Review

Defining as well as classifying compounds is varied and complex. A compound is the product of the fixed juxtaposition of two free words that otherwise occur independently (Adams, 1973). Free words like *home* and *work* are combined to create the compound word *homework*. This definition provides a useful description of compounds; however, it is not comprehensive, as there are some compounds that combine more than two lexical units, as in *long-term loan*. Compounds have a strong

tendency to be binary, with precisely two free morphemes in juxtaposition. Compounds with more than two roots are often broken down into a sequence of binary compounds. The word in *long-term loan* is a binary compound that comprises *long-term* and *loan*, and then the first element of the compound, *long-term*, can be classified into *long* and *term*. Notwithstanding the binary nature of compounds and the prevalence of two-morpheme compounds in English, a compound, according to a more exhaustive definition, is a word composed of two or more lexical morphemes or lexemes (e.g., Langacker, 1973; Bauer, 1998).

Based on the definitions provided above, one can conclude that all examples of compounds comprise free morphemes. However, there are certain English compounds in which at least one element is bound, namely neoclassical compounds. For example, the compound *cranberry* comprises the bound morpheme *cran* and the free morpheme *berry*. Halpern (2000) acknowledges that the constituents of a compound are not always free though he does not incorporate this in his definition of compounds.

Based on semantic criteria, Bauer (1983) divides compound nouns into four categories. Endocentric compounds are those that denote hyponyms of the compounds' head components. The compound *housework*, for example, is a kind of *work*. Exocentric compounds (also known as Bahuvrihi in Sanskrit) are compounds that are not hyponyms of the grammatical head. The compound *pickpocket*, for example, is not a type of pocket. In these compounds, the lack of a specified semantic head is frequently perceived as metaphorical or synecdochic. Appositional compounds are those in which both components have the possibility to be the head of the compound. For instance, *player-coach* is a hyponym of both *player* and *coach*. Copulative compounds (also known as dvandva in Sanskrit) are those that are not hyponyms of either constituents and refer to separate entities that combine to produce the entity represented by the compound, such as *Rank-Hovis* (Bauer, 1983).

It is claimed that no genus-species compounds, such as *humanman and *placemoor, should exist because the determining element is implicit in the head component. However, there are several exceptions to this rule. Genus-species compounds (also called clarifying or classifying compounds by Gusmani (1973), cited in Grzega, 2002), though they may appear redundant, are ubiquitous in English (Bauer, 1983). *Cod fish*, *beech tree*, and *boy child* are few examples. Bauer's (1983) categorization is of a structural character as the concept of headedness is present in all classifications, whether it is present, missing, or shared in the compound.

Additional forms of compounds are reduplication compounds and synthetic (or linguistic) compounds, according to Fabb (1998). Reduplication compounds, like *hush-hush* (secret) and *tick-tock* (clock sound), comprise identical (or nearly identical) constituents. Synthetic (or linguistic) compounds are those in which the head constituents are derived words that include verbs and one of a set of affixes (usually -er, -ing and -en), as in *family planning* and *shoemaker*. Compound categorization by Fabb (1998) is basically a structural classification of compounds.

A different classification of compounds by Adams (1973) employs both grammar and meaning. The primary focus of this study is on the link between the elements of compounds, utilising grammatical relations in some examples and semantic relations in others. She divides compounds into eleven categories: Subject-Verb, Verb-Object, Appositional ('B that acts as, has the function of A', 'B of which A is a particular instance', 'B is an A'), Associative ('B is part of A', 'B belongs to A', 'B is typically associated with A', 'B is produced or derived from A'), Instrumental ('B which prevents or cures against A', 'B which is the means of preserving A' and 'B which causes or promotes A', Locative ('A is a place where or a time when B is or happens, Resemblance ('B which is in the form of, has the physical features of, A' 'B which reminds one of A', Composition / Form / Contents (one element specifies the other in relation to some concrete feature), Adjective-Noun, Names, Other (some

cases do not belong to any of the provided classes). Adams's (1973) classification is elaborate; however, it is not exhaustive, as certain compounds remain unclassified.

As mentioned earlier, exocentric compounds attracted less attention than regular compounds for several reasons. First, the approaches that linguists adopted in categorising compounds are classical. Under this approach, to be included in a category, an entity must meet all the necessary and adequate criteria. To put it another way, an entity that does not meet all the necessary and adequate conditions will not be classified as a member of a category. Exocentric compounds rarely follow morphogrammatical rules, and so they are non-members of the category of morphology. Second, the previous approaches to word-formation studied word-formation processes in distinct linguistic disciplines as research on linguistics breaks down into different fields, notably syntax, morphology, phonology, and semantics. A key limitation of this approach, despite its invaluable contributions, is that it does not address the interfaces among the diverse characteristics of language. Within this framework, studies on compounds focused on only their syntactic and morphological constituents. This conflicts with the natural formation of a neologism, which is the product of a uniform procedure in the mind of a speaker that a hearer simultaneously understands regardless of which component is prevalent. Finding the structure rules shared by all the different facets of language is pivotal (Evans & Green, 2006). Last but not least, because of the transformationalist/generativist framework's influence, syntax predominated in all studies on the diverse language phenomena. As Benczes (2006) points out, this approach neglected morphology for a long period. Such neglect resulted from the fact that post-syntactic structures linguistics witnessed the omnipresence of phonology and syntax, causing morphology to be lost somewhere in the middle (Aronoff, 1976).

Morphological devices cannot be subject to the same grammatical constraints, as there are no rules without exceptions. However, such exceptions should not be disregarded because they may be constrained by other linguistic rules that need research exploration. If such exceptional cases are just kept as listemes in the lexicon, many linguistic phenomena will go unexplored. A reasonable and thorough analysis of the English language must also account for exceptional categories and explicate their existence (Bauer & Renouf, 2001). Syntactic and grammatical rules cannot account for all the semantic features of lexical items. Compounding, unlike affixation, is not constrained by grammatical constraints, except for some fundamental semantic needs and extra-grammatical pragmatic concerns. The nameability criterion, for example, emphasizes that a lexical item must not only represent something actual to the speaker but also something that can be named (Bauer, 1983).

No one can deny that endocentric compounds make up the vast bulk of compounds, making them archetypal instances of compounds. Less archetypal instances like exocentric compounds also contribute to the English vocabulary and should not be overlooked. The marginalisation of exocentric compounds, as well as other key linguistic phenomena, reflects the previous approaches' failure to explicate them under the guise that they deviate from the morphosyntactic rules. Being exceptional for grammatical rules does not mean that exocentric compounds are unanalysable or do not fit into systematic patterns. It was not possible for prior approaches, such as descriptivist, transformationalist and generativist approaches, to map the semantics of metaphorical and/or metonymical compounds systematically (Benczes, 2006) since exocentric compounds are founded on metaphor and metonymy.

Conceptual metaphor and conceptual metonymy usually motivate the creation of exocentric compounds. These cognitive mechanisms were not taken into consideration until cognitive linguistics discovered that metaphor is widespread in language, thinking, and action. According to Lakoff and Johnson (2008), the common conceptual system of humans is metaphorical in terms of thinking and action. Before such discovery, as well as redefinition of, metaphor's function in human cognition and language, it was traditionally considered being merely a rhetorical tactic used for persuasion. Metaphor, according to classical definitions, can be described as a linguistic phenomenon, an artistic

or rhetorical device, based on two entities that are like one another, a conscious utilisation of words, and something that we can do without (Benczes, 2006). Priority should therefore be given to the conceptual and semantic aspects, which serve as the foundation for the formation of compounds. Instead of relegating "non-conforming" data to the "exception to the rule" pile, an approach that views endocentric and exocentric compounds as prototypical subcategories or members of the category compounds, acknowledges the fuzzy border between them, and emphasizes their conceptual foundation may provide a better insight into the speaker's creation of compounds in general and justification for all data (cf. Lahlou & Ho-Abdullah, 2021).

According to cognitive linguistics, metaphor and metonymy are used not only in literature but also in everyday speech. Metaphor is a conventional association of one domain with another domain. It is conceptual since metaphor motivation occurs at the conceptual level (Lakoff & Johnson, 2008). In the expression *He just sails through life*, for instance, life is understood in terms of a journey. The life of a traveller is a journey, as is the life of a person leading a life. The metaphorical mappings between the way of travelling and that of living, a traveller's destination and a person's life goals, and the way physical obstacles and life difficulties are formed (Kövecses, 2015). The underlying mechanism by which humans comprehend abstract concepts and perform abstract reasoning is metaphor (Lakoff, 1993). It is required in verbal communication because it allows people to express and think about abstract concepts in relation to concrete concepts (e.g., Kövecses, 2010, 2015; Lahlou, 2018; Lahlou & Rahim, 2021; Lakoff & Johnson, 2008).

The second conceptual mechanism, which motivates the forming of exocentric compounds is metonymy. Metonymy occurs when an entity refers to another entity to which it is connected (Lakoff & Johnson, 2008). In *The kettle's boiling*, for example, to use Lakoff and Johnson's example, *the kettle* refers to the water in the kettle. The kettle (the vehicle), in this context, provides mental access to the water in the kettle (the target) within the same domain or idealized cognitive model (ICM), to employ Radden and Kovecses' (1999) definition.

Despite the difference between conceptual metaphor and conceptual metonymy in mapping as the first takes place between two different conceptual domains, while the second occurs within the same conceptual domain, across conceptual domains, they share many characteristics. Both of them are conceptual in nature and can be described as mapping processes. Both can be conventionalized, that is, automatic, unconscious, effortless, generally established as a model of thought, and ways of expanding a language's resources (Lakoff & Turner, 1989).

3. Methodology

The corpus of the current study consists of one hundred exocentric compounds randomly selected from prior studies on compounds, namely Bloomfield (1933), Bauer (1983), Stekauer (2001), Carstairs-McCarthy (2002), Libben et al., (2003), Aronoff and Fudeman (2011), and Hamawand (2011). Compiling data from earlier research will aid in the continuation of the linguistic debate and embody the notion of exocentric compounds in the literature.

All the examples of exocentric compounds in the data set are compound nouns that consist of two constituents. The internal grammatical structure of the exocentric compounds comprises Noun + Noun, Adjective + Noun, and Verb + Noun. Table 1 displays these compound formations with examples.

Table 1. Internal structures of the compounds in the selected data set

Compound internal structure	Example
Noun + Noun	birdbrain
Adjective + Noun	blackhead
Verb + Noun	skinflint

To identify the frequency of the selected collection of exocentric compounds, the current study employs Sketch Engine, a web-based Corpus Query System (CQS), through which the selected corpora are accessed, namely the *British National Corpus* (BNC) and the *English Web 2020* (enTenTen20). The BNC is a huge representative collection of texts (90%) and dialogues (10%) from a wide range of sources from the late twentieth century, totalling 96,135,000 words. However, the BNC contains no texts beyond 1994 (Thomas, 2017); therefore, it was pertinent to employ a more recent and even larger corpus, the enTenTen20, to attain more exhaustive up-to-date data on exocentric compounds. The enTenTen20 comprises 38,149,437,411 words, collected from diverse Internet sources between 2019 and 2021. In comparison with the BNC, which consists of samples of English written and spoken language in Britain, the enTenTen20 comprises English texts from the Web and so from different parts of the world, including Britain, the USA, and Australia.

To identify the frequency of the selected exocentric compounds, a Simple Query on each exocentric compound was performed in the selected corpora separately, displaying all its concordances. At the top of the concordance page, the result details, including the raw frequency of hits and the number of hits per million tokens, appear. The number of hits per million tokens, among others, is crucial to normalize the frequencies to a common base. Normalization is typically used to compare corpora or subcorpora of varying sizes accurately and fairly (McEnery & Hardie, 2012).

In addition, in corpus linguistics, normalizing the frequencies to one million is a common baseline (Brezina, 2018). However, because of the diversity in the orthography of compounds, the authors manually normalised some of the occurrence counts. A compound can be written as a single word, for example, *sawbones* (a surgeon), as hyphenated words, for example, *jail-bird* (a person who is/has been in prison), or as two separate words, for example, *loony bin* (a psychiatric hospital). Thus, when a compound has two forms, as with *jail-bird*, which can also be written as *jailbird*, Simple Query searches for both forms are performed, and then the results are manually normalized. To normalize the frequencies, the following formula was employed.

Table 2. Relative frequency calculation

$$Relative\ frequency = \frac{absolute\ frequency}{number\ of\ tokens\ in\ corpus} \times basis\ for\ normalization$$

Adapted from Brezina (2018)

Thus, the authors computed the relative frequency of the exocentric compounds *jail-bird* (1 hit) and *jailbird* (18 hits) in the BNC as follows:

Table 3. Relative frequency calculation of jail-bird/jailbird

$$\text{Relative frequency } jail-bird/\ jailbird = \frac{19}{112{,}345{,}722} \times\ 1{,}000000\ =\ 0.17$$

On the whole, the raw frequency (RF) of *jail-bird*/*jailbird* is 19, while its normalized frequency (NF) is 0.17.

To explore the extent to which conceptual metaphor and/or conceptual metonymy motivate the creation of exocentric compounds, the selected data set was manually investigated. Some exocentric compounds are metonymically derived. The meaning of the exocentric compound *cut-throat* (or *cutthroat*), a combination of the verb *cut* and noun *throat*, for example, is metonymically projected to the meaning of 'murderer'. In this example, *cut-throat* refers to a person who cuts throats; thus, the source domain ACTION (i.e., an action of cutting throats) is mapped onto the target domain AGENT (i.e., an agent cutting throats) in the same domain. The creation of this exocentric compound is then motivated by ACTION FOR AGENT conceptual metonymy.

Some exocentric compounds are formed based on conceptual metaphor. For instance, the compound *rug rat* (or *rug-rat*), 'a child or toddler', is the outcome of a metaphor-based semantic extension. The child is identified with some of a rat's attributes, notably crawling and making a mess. The source domain RODENT (rat) is metaphorically mapped onto the target domain HUMAN (toddler). The formation of this compound is thus motivated by HUMAN IS RODENT conceptual metaphor. This is a metaphor because it is a conventional association of one domain with another domain and conceptual, since the motivation for metaphor occurs at the conceptual level (Lakoff & Johnson, 2008).

Both conceptual metonymy and conceptual metaphor motivate the creation of some other exocentric compounds. A good example is the compound noun *lion heart*, 'a very brave person'. First, the source domain PART (i.e., *heart*) stands for (or gets access to, to use the cognitive terminology) the target domain WHOLE (i.e., a person). Therefore, the formation of this compound is motivated by PART FOR WHOLE conceptual metonymy. Second, the person is associated with one of a lion's qualities, namely bravery, and therefore the source domain LION is mapped onto the target domain HUMAN. This metaphor is based on cultural knowledge that believes that a lion is brave, rather than physical resemblance. This form of metaphor is known as resemblance metaphors, according to Grady (1999), who concluded that metaphors are grounded in experiential correlations rather than similarities (cited in Lakoff & Johnson, 2008).

4. Results and Discussion

The data collected from the selected corpora showed that exocentric compounds survived to the present. The overall results on exocentric compound frequency in the BNC and the enTenTen2020 are reported in Table 4.

Table 4. The raw and normalized frequencies of the selected exocentric compounds

Exocentric compounds	BNC RF & NF	enTenTen20 RF & NF	Exocentric compounds	BNC RF & NF	enTenTen20 RF & NF
birdbrain	1 (0.01)	966 (0.02)	humbug	89 (0.79)	16,824 (0.37)
blackhead	21 (0.19)	25,329 (0.56)	hunchback	46 (0.41)	14,769 (0.33)
blackbird	299 (2.66)	53,770 (1.2)	Jail-bird	20 (0.18)	3,411 (0.08)
blockhead	7 (0.06)	7,467 (0.17)	Killjoy	28 (0.25)	7,521 (0.17)
bluebell	135 (1.2)	29,153 (0.65)	lazybones	2 (0.02)	819 (0.02)
blue-coat	-	230 (0.01)	lion heart	1 (0.01)	2,389 (0.05)
bluegill	1 (0.01)	15,936 (0.35)	Longnose	1 (0.01)	1,910 (0.04)
blue pencil	7 (0.06)	2,416 (0.05)	loony bin	8 (0.07)	1,520 (0.03)
blue-stocking	15 (0.13)	3,765 (0.08)	loudmouth	18 (0.16)	10,224 (0.23)
bonehead	3 (0.03)	8,969 (0.2)	lowlife	52 (0.46)	13,880 (0.31)
bootblack	7 (0.06)	2,117 (0.05)	makeshift	270 (2.4)	86,850 (1.93)
brass hat	4 (0.04)	734 (0.02)	mainstream	1,278 (11.38)	656,537 (14.6)
breakfast	4,305 (38.32)	1,550,568 (34.48)	oddball	62 (0.55)	31,053 (0.69)
breakwater	92 (0.82)	33,308 (0.74)	paleface	3 (0.03)	1,985 (0.04)

buttercup	75 (0.67)	21,268 (0.47)	paperback	562 (5)	314,444 (6.99)
card shark	2 (0.02)	3,012 (0.07)	passport	1,047 (9.32)	521,927 (11.61)
catchfly	-	699 (0.02)	pickpocket	51 (0.45)	16,807 (0.37)
climbrock	-	-	pronghorn	1 (0.01)	9,581 (0.21)
copperhead	3 (0.03)	12,105 (0.27)	ragtime	38 (0.34)	27,912 (0.62)
cottonmouth	2 (0.02)	3,781 (0.08)	razorback	1 (0.01)	20,160 (0.45)
cottontail	2 (0.02)	10,951 (0.24)	redbreast	13 (0.12)	2,064 (0.05)
cut-throat	95 (0.85)	43,244 (0.96)	redhead	107 (0.95)	60,969 (1.36)
daredevil	47 (0.42)	59,544 (1.32)	redskin	20 (0.18)	79,573 (1.77)
deadline	1,104 (9.83)	1,079,119 (24)	rubberneck	60 (0.53)	479 (0.01)
dimwit	8 (0.07)	6,883 (0.15)	rugrat	-	2,394 (0.05)
dingbat	9 (0.08)	4,551 (0.1)	sabertooth	-	5,018 (0.11)
doughnut	123 (1.09)	71,255 (1.58)	sawbones	4 (0.04)	1,402 (0.03)
egghead	23 (0.2)	7,501 (0.17)	scarecrow	67 (0.6)	45,701 (1.02)
fathead	4 (0.04)	5,585 (0.12)	scatterbrain	2 (0.02)	1237 (0.04)
featherbrain	1 (0.01)	138 (0.01)	shoehorn	10 (0.09)	7,338 (0.16)
foxglove	67 (0.6)	10,614 (0.24)	short-horn	109 (0.97)	5,398 (0.12)
Figurehead	98 (0.87)	30,336 (0.67)	skinflint	16 (0.14)	1,847 (0.04)
five-finger	2 (0.02)	2,194 (0.04)	skyscraper	123 (1.13)	103,973 (2.31)
fleabag	2 (0.02)	4,281 (0.1)	slowpoke	--	3,844 (0.09)
Funny farm	16 (0.14)	1,442 (0.03)	snap-dragon	10 (0.09)	59,802 (1.33)
goldeneye	23 (0.2)	15,826 (0.35)	sourpuss	8 (0.07)	1,999 (0.04)
goldenrod	11 (0.1)	11,276 (0.25)	spoil-sport	40 (0.36)	3,823 (0.09)
greenbelt	555 (4.94)	36,699 (0.82)	spoonbill	12 (0.11)	8,328 (0.19)
greenhorn	4 (0.04)	9,005 (0.2)	staircase	1025 (9.12)	213,360 (4.74)
greybeard/ graybeard	14 (0.12)	4,874 (0.11)	stalemate	202 (1.8)	53,262 (1.18)
hallmark	354 (3.15)	174,882 (3.89)	straightedge	2 (0.02)	3,943 (0.09)
hammerhead	8 (0.07)	23,162 (0.52)	strawberry	634 (5.64)	313,399 (6.97)
hangdog	21 (0.18)	1,977 (0.04)	swallowtail	14 (0.13)	13,090 (0.29)
hardback	214 (1.9)	60,665 (1.35)	telltale	205 (1.82)	54,225 (1.21)
hardtop	6 (0.05)	21,909 (0.49)	turncoat	14 (0.12)	7,065 (0.16)
head-hunter	211 (1.88)	15,984 (0.36)	undercoat	11,993 (0.27)	55 (0.49)
heatwave	49 (0.44)	31,216 (0.69)	wagtail	41 (0.36)	13,725 (0.31)
highbrow	76 (0.68)	10,292 (0.23)	walkman	178 (1.58)	16,069 (0.36)
hogwash	9 (0.08)	10,091 (0.22)	windfall	198 (1,76)	52,079 (1.16)
Hot dog	90 (0.8)	118,794 (2.64)	whitecap	2 (0.02)	17,341 (0.39)

As shown in Table 4, the normalized frequency of the exocentric compounds ranges from 0.01 (pmw) to 38.32 (pmw) in the BNC and 0.01 (pmw) to 34.48 (pmw) in the enTenTen20. There are seven hapax legomena (or hapaxes), words that occur only once in the corpus, in the BNC. There are also nine dis legomena, words appearing only twice. This may be attributed to linguists' use of intuition instead of corpora in selecting examples given that the data set under study is collected from previous works on exocentric compounds. This can be supported by the use of the compound *climbrock*, which is not available in both corpora. Hapaxes, however, as postulated by Zipf's law, are prevalent and play an important role in corpus-based studies. The relative frequency of hapaxes in a corpus, for example, can be used to examine how diverse the vocabulary is (Baker, Hardie & McEnery, 2006). According to Kornai (2008), in large corpora, approximately 40% to 60% of all lexical units appear only once, and an additional 10% to 15% appear only twice. In addition, several exocentric compounds were coined in American English like *bluecoat*, 'a person wearing a blue coat, such as a soldier', *rugrat*, 'a child', *slowpoke*, 'a person that acts too slowly'. Therefore, their usage is less in the BNC.

Despite the frequency differences between the BNC and the enTenTen20, the data show exocentric compounds can survive to the present. According to the BNC, they were employed in British English

in the late twentieth century and in all varieties of English, including American English and Australian English, in the twenty-first century according to the enTenTen20. This shows their significant function in the English vocabulary.

Another important finding was that exocentric compounds also typically contributed to the English vocabulary by deriving other lexical words, namely verbs, adjective and adverbs, and even other compound nouns. The adjective *mainstream* and verb *to mainstream*, for example, are derived from the compound noun *mainstream*, 'the prevailing current of thought'. This compound noun also generated the compound noun *mainstream media*, 'conventional newspapers, television and other news sources that most people know about and regard as reliable'. More importantly, numerous exocentric compounds become polysemous, contributing new meanings and so enriching the English vocabulary. Consider, for example, the compound noun *passport*, originating from Old French with the meaning of 'permission to depart from a port or harbour'. The original meaning became obsolete, and new meanings developed, namely, 'any authorization to pass or go somewhere', 'a document issued by a country to a citizen allowing that person to travel abroad and re-enter the home country', and 'recommendation, any quality or characteristic that gains a person a favourable reception or acceptance or admission'.

The semantic analysis of the exocentric compound data set shows how prominent conceptual metonymy and conceptual metaphor, the major cognitive mechanisms, are in creating exocentric compounds. Figure 1 outlines the categorization of the cognitive mechanisms motivating the coining of the exocentric compounds under study.

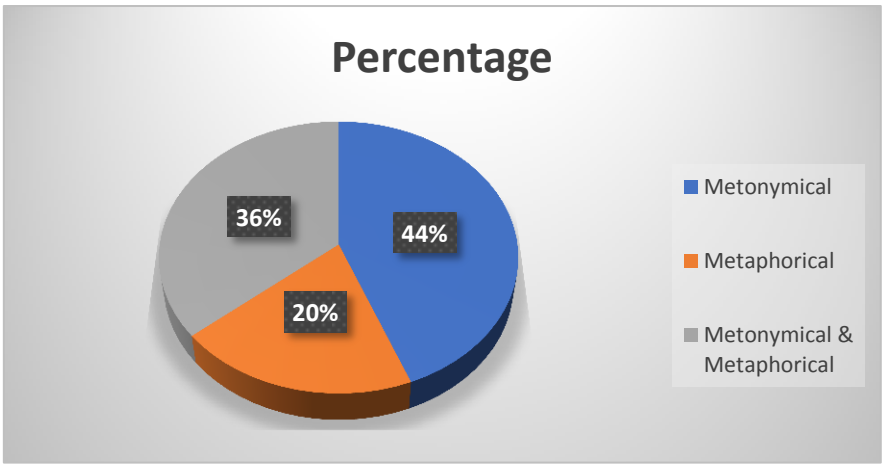

Figure 1. Cognitive mechanism motivation percentage

As shown in Figure 1, conceptual metonymy is more active in deriving exocentric compounds than conceptual metaphor, accounting for over a third (36%) of the cognitive motivation as a single cognitive factor and over a third (44%) of the cognitive motivation when combined with conceptual metaphor. This finding is significant because it shows the importance of conceptual metonymy in the formation of exocentric compounds. This adds to the evidence for the importance of conceptual metonymy in extending the resources of the English language, as well as its comparability to conceptual metaphor. According to several studies, metonymy is as common in language and thought as metaphor (Panther & Thornburg, 2003).

The second most important finding from the data set is that both conceptualization processes motivate 44% of the exocentric compounds. This emphasizes how conceptual metonymy and conceptual metaphor are related and adds to the list of postulated common features shared by these conceptualization processes.

5. Conclusion

The current study looked into the frequency of exocentric compounds employed by linguists in the literature, as well as the extent to which conceptual metaphor and/or conceptual metonymy motivate exocentric compound formation. The findings show that exocentric compounds not only provide the English language with creative lexical units to fill vocabulary gaps and allow English users to more accurately convey their intended meaning, but they are also viable lexical items that generate lexical words like adjectives through productive word-formation processes, as well as semantic extensions. The study findings also show that conceptual metonymy is highly activated in the formation of exocentric compounds, and that it is comparable to conceptual metaphor. Given the study objectives, the data set investigated in the current study is representative of the examples used by researchers in the literature. Thus, further research on exocentric compounds, employing a bigger corpus, is desirable to gain more insight into it.

Acknowledgement

This research is partly funded by Universiti Sains Malaysia Short Term Research Grant (No. 304/PHUMANITI/6315486).

References

Adams, V. (1973). *An introduction to modern English word-formation*. London: Longman.

Adams, V. (2014). *Complex Words in English*. New York: Routledge.

Allen, M. R. (1979). *Morphological investigations*. Doctoral dissertation, University of Connecticut.

Aronoff, M. (1976). Word formation in generative grammar. *Linguistic Inquiry Monographs Cambridge, Mass*, (1), 1-134.

Aronoff, M., & Fudeman, K. (2011). What is morphology? (Vol. 8). John Wiley & Sons.

Baker, P., Hardie, A., & McEnery, T. (2006). *A Glossary of Corpus Linguistics*. Edinburgh: Edinburgh University Press.

Bauer, L. (1983). *English word-formation*. London: Cambridge.

Bauer, L. (1998). Is there a class of neoclassical compounds, and if so is it productive? *Linguistics 36*(3), 403–422.

Bauer, L., & Renouf, A. (2001). A corpus-based study of compounding in English. *Journal of English Linguistics, 29*(2), 101–123.

Benczes, R. (2006). *Creative compounding in English*. Amsterdam and Philadelphia: John Benjamins Publishing Company.

Bloomfield, L. (1933). *Language*. London: George Allen and Unwin.

Brezina, V. (2018). *Statistics in corpus linguistics: A practical guide*. Cambridge: Cambridge University Press.

Carstairs-McCarthy, A. (2002). *An Introduction to English Morphology: Words and Their Structure*. Edinburgh: Edinburgh University Press.

Evans, V., & Green, M. (2006). *Cognitive Linguistics: An Introduction*. Edinburgh: Edinburgh University Press Ltd.

Fabb, N. (1998). Compounding. In A. Spencer & A. Zwicky (Eds.), *The Handbook of Morphology* (pp. 66-83). Oxford: Blackwell.

Grzega, J. (2002). Some thoughts on a cognitive onomasiological approach to word formation with special reference to English. *Onomasiology Online*, *3*, 55-81.

Halpern, J. (2000). *Is English Segmentation Trivial?* Technical report, CJK Dictionary Institute.

Hamawand, Z. (2011). *Morphology in English: Word formation in cognitive grammar*. London: Continuum.

Katamba, F. (2006). *Morphology* (2nd ed.). New York: Palgrave Macmillan.

Kornai, A. (2008). *Mathematical linguistics*. Berlin: Springer.

Kövecses, Z. (2010). *Metaphor: A practical introduction*. Oxford University Press.

Kövecses, Z. (2015). *Where metaphors come from: Reconsidering context in metaphor*. New York: Oxford University Press.

Lahlou, H. (2018). The conceptualisation of science terminology: A cognitive linguistic analysis of the categories ELECTRICITY and LIGHT in Arabic. *International Journal of Humanities and Social Science Research*, *4*(2), 75-80.

Lahlou, H., & Abdullah, I. H. (2021). The fine line between compounds and portmanteau words in English: A prototypical analysis. *Journal of Language and Linguistic Studies, 17*(4), 1684-1694. https://doi.org/10.52462/jlls.123

Lahlou, H., & Rahim, H. A. (2022). Conceptual metaphors in North African French-speaking news discourse about COVID-19. *Indonesian Journal of Applied Linguistics*, *11*(3), 589-600. https://doi.org/10.17509/ijal.v11i3.35949

Lakoff, G., & Turner, M. (1989). *More Than Cool Reason: A Field Guide to Poetic Metaphor*. Chicago: University of Chicago Press.

Lakoff, G. (1993). The contemporary theory of metaphor. In A., Ortony (Ed.), *Metaphor and Thought* (pp. 202-251). New York: Cambridge.

Lakoff, G., & Johnson, M. (2008). *Metaphors We Live by*. Chicago and London: The University of Chicago Press. (Original work published 1980)

Langacker, R. W. (1973). *Language and its structure: some fundamental linguistic concepts* (2nd ed.). New York: Harcourt Brace Jovanovich.

Libben, G., Gibson, M., Yoon, Y. B., & Sandra, D. (2003). Compound fracture: The role of semantic transparency and morphological headedness. *Brain and language*, *84*(1), 50-64.

McEnery, T. & Hardie, A. (2012). *Corpus Linguistics: Method, Theory and Practice*. Cambridge: Cambridge University Press.

Panther, K. U., & Thornburg, L. (2003). The EFFECT FOR CAUSE metonymy in English grammar. In A. Barcelona (Ed.), *Metaphor and Metonymy at the Crossroads: A Cognitive Perspective* (pp. 215-232). Berlin: Mouton de Gruyter.

Radden, G., & Kövecses, Z. (1999). Towards a theory of metonymy. *Metonymy in language and thought*, *4*, 17-60.

Spencer, A. (1991). *Morphological Theory: An Introduction to Word Structure in Generative Grammar*. Oxford: Wiley-Blackwell.

Stekauer, P. (2001). Fundamental principles of an onomasiological theory of English word-formation. *Onomasiology Online*, *2*, 1-42.

Szymanek, B. (2005). The latest trends in English word-formation. In P. Štekauer, & R. Lieber (Eds.), *Handbook of Word-Formation* (pp. 429-448). Springer, Dordrecht.

Ten Hacken, P. (2000). Derivation and compounding. In Booij, G., Lehmann, C., & Mugdan, J (Eds.), *Morphologie / Morpho*logy (pp. 349–360). Berlin: Walter de Gruyter.

Williams, E. (1981). On the notions "Lexically related" and "Head of a word". *Linguistic inquiry*, *12*(2), 245-274.

AUTHOR BIODATA

Hicham Lahlou (PhD) is a Senior Lecturer in the English Language Studies Department of the School of Humanities at Universiti Sains Malaysia. His research interests include cognitive linguistics, semantics, morphology and corpus linguistics.

Imran Ho-Abdullah (PhD) is a Professor at the School of Language Studies and Linguistics, Faculty of Social Sciences and Humanities, Universiti Kebangsaan Malaysia. Currently he is Deputy Vice-Chancellor (Industry and Community Partnership) at Universiti Kebangsaan Malaysia. His areas of interest include corpus linguistics, cognitive linguistics, and semantics.