Univerzita Karlova
Filozofická fakulta
Ústav anglického jazyka a didaktiky
filologie – anglický jazyk

Kateřina Vašků

Lexical idioms in English

Lexikální idiomy v angličtině

Disertační práce

Vedoucí práce: Prof. PhDr. Aleš Klégr

2018

Abstrakt

Podle standardní definice se frazeologie zabývá víceslovnými lexikálními jednotkami, tzn. kombinacemi slov. Hlasy volající po tom, že i komplexní slova složená ze dvou či více významových jednotek mohou mít status (lexikálních) frazémů/idiomů, zvláště je-li jejich význam nekompozicionální, jsou stále dosti izolované, a to i přes to, že lingvistická literatura se hemží zmínkami o idiomatických kompozitech a derivátech (Kap. 3). Zdá se, že jediné systematické pojednání o lexikálních idiomech podává František Čermák (2007), který se zaměřuje především na lexikální idiomy v češtině. Cílem této práce je proto prozkoumat situaci v angličtině a pokusit se vytvořit nosnou definici a zejména kritéria pro odlišení lexikálních idiomů od ostatních komplexních lexémů a nastínit hlavní typy těchto idiomů v angličtině. Po úvodu (Kap. 1) a seznámení se současnými proudy ve frazeologii a relevantními poznatky o frazeologických jednotkách a jejich rysech (Kap. 2), referuje práce o Čermákově teorii lexikálních idiomů a kvantitativní studii, kterou jeho teorie inspirovala (Kap. 4). Jádrem práce je analýza dvou vzorků. První byl vybrán z BNC a představuje náhodný výběr 1000 jednoslovných lemmat. Sloužil jako testovací vzorek nejen pro odlišení simplexních lemmat od komplexních, ale zejména pro zjišťování potenciálních lexikálních idiomů, a tím k upřesnění jejich výchozí definice s důrazem na sémantické anomálie. Druhý vzorek vytvořený na základě Oxford English Dictionary se skládá z 500 preselektovaných lexémů vyznačujících se sémantickou anomálií (někdy spojenou s dalším typem anomálie), Jejich rozbor ukázal, že představují škálu lexikálních idiomů od centrálních až po periferní různého typu z hlediska výskytu a kombinace anomálií (Kap. 5 a 6). Závěry práce (Kap. 7) do značné míry podporují Čermákova zjištění pro češtinu, nicméně poukazují i na specifické rysy anglických lexikálních idiomů. Celkově vyznívají ve prospěch uznání kategorie lexikálních idiomů v angličtině a jejich zařazení do frazeologie jako legitimního předmětu zkoumání s tím, že jde o kategorii hraniční, která plynule přechází v lexémy kompozicionální neidiomatické.

Klíčová slova: frazeologie, jednoslovné lexémy, lexikální idiomy, idiomatičnost, nekompozicionalita, sémantická anomálie

Abstract

According to the standard definition phraseology deals with multi-word lexical units, i.e. word combinations. Voices claiming that even complex words composed of two or more meaningful units may qualify for the status of (lexical) phrasemes/idioms, especially when their meaning is non-compositional, are still very isolated, in spite of the fact that linguistic literature is teeming with references to idiomatic compounds and derivatives (Chap. 3). In fact, the only systematic treatment of lexical idioms seems to be that offered by Čermák (2007), who focuses primarily on lexical idioms in Czech. The aim of the thesis is therefore to explore the situation in English and attempt to develop a useful definition of, and especially criteria for, distinguishing lexical idioms from other complex lexemes and provide an outline of the main types of lexical idioms obtaining in English. After an introduction (Chap. 1) and the presentation of state-of-the-art approaches to phraseology and the relevant information about phraseological units and their features (Chap. 2), the thesis reviews Čermák's theory of lexical idioms which inspired their quantitative study in Czech (Chap. 4). The core part is the analysis of two samples. The first one, gathered from the BNC, includes a random selection of 1000 single-word lemmas and served as a testing ground for not only separating simple lexemes from complex ones, but especially for identifying potential lexical idioms and thus for recasting the initial definition of lexical idioms with an emphasis on semantic anomaly. The second sample, based on the OED, consists of 500 preselected lexemes exhibiting semantic anomalies (sometimes combined with other types of anomaly) whose analysis revealed them to range from central to peripheral lexical idioms and display a variety of types in terms of anomaly combinations (Chap. 5 and 6). The conclusions of the thesis (Chap. 7) largely support Čermák's findings about Czech lexical idioms, yet pointing out specific features of English lexical idioms, and argue for the recognition of the category of lexical idioms in English and their legitimate inclusion within the scope of phraseology, although as a borderline category shading off into compositional non-idiomatic lexemes.

Keywords: phraseology, single-word lexemes, lexical idioms, idiomaticity, non-compositionality, semantic anomaly

Prohlašuji, že jsem disertační práci napsala samostatně s využitím pouze uvedených a řádně citovaných pramenů a literatury a že práce nebyla využita v rámci jiného vysokoškolského studia či k získání jiného nebo stejného titulu.

V Praze, dne 12. 7. 2018 **……..……**

Největší poděkování patří mému školiteli Prof. PhDr. Aleši Klégrovi za velkou trpělivost, podporu a motivaci a také za obětavou pomoc a cenné rady, které mi při řešení práce velmi pomohly. Děkuji také všem přátelům z Ústavu anglického jazyka a didaktiky za důvěru, kterou ve mě vložili. V neposlední řadě děkuji svému muži a dětem za to, že pro mé úsilí měli pochopení.

Contents

Abbreviations

ADJ	adjective
ADV	adverb
BNC	British National Corpus
CD	Collins Dictionary online
CL	complex lexemes
COCA	Corpus of Contemporary American English
ex	example
exx	examples
N	noun
NN	noun + noun compound
OALD	Oxford Advanced Learner's Dictionary
OED	Oxford English Dictionary
P	particle
PI	phraseme/idiom
PP	past participle
Prep	preposition
V	verb
WF	word-formation

List of tables

List of figures

1. Introduction

The subject of the present study is the analysis of lexical idioms in English, i.e. single-word complex lexemes which exhibit similarities to multi-word idioms recognized traditionally in phraseology. The main question inspected in the study is whether it is plausible to study anomalous combinatorial relations below the level of the word within and by means of phraseology.

Describing such items as idioms is not generally accepted yet, inasmuch as idioms are canonically thought to be multi-word or polylexical expressions (cf. Granger & Paquot, 2008). However, some recent trends in linguistics (especially within the field of cognitive linguistics, cf. Onysko, Michel, 2010) emphasize that the borderlines between traditional levels of language study, esp. between morphology and syntax/phraseology are blurred and fuzzy. The present study is an attempt to argue for the possibility of describing and analysing lexical idioms by similar criteria as idiomatic multi-word expressions.

The theoretical part is represented by Chapters 2, 3 and 4. Chapter 2 provides an overview of the traditional and more recent approaches to phraseology, presents structural classifications of phrasemes in the work of some of the most influential phraseologists within the phraseological approach, and elaborates also on semantic aspects of complex units, i.e. compositionality and idiomaticity. Chapter 3 illustrates by evidence from the literature that single-word lexemes have been relatively widely analysed in phraseological terms in the non-phraseological literature for quite some time, although the first systematic classification of idioms which includes also single-word lexemes is that of F. Čermák (1982, 2007a, 2007b). His description of lexical idioms is presented in Chapter 4.

The empirical part presents a provisional definition of lexical idioms – drawing on the definitions of Czech lexical idioms offered by Čermák (2007a, 2007b) and Klötzerová (1997, 1998) – which is then tested on a sample of 1000 randomized lemmas from BNC (see Chapter 3 for more detail). The aim of their analysis is to examine all anomalous aspects of morpheme combinations (both formal and semantic). Since the definitions by Čermák and Klötzerová are primarily used for Czech instances of lexical idioms, it was expected that some revisions

would have to be made to adapt the definition to English (which is different typologically and whose vocabulary is stratified differently due to its historical development). Based on the analysis of the BNC sample, a modified definition of lexical idioms is proposed. This definition is then used to extract a sample of 500 lexemes from OED (see Chapter 3 for more detail) which can be described as lexical idioms. These are then classified according to the type of anomaly they exhibit, their formal structure and their pragmatic function, in order to find similarities to multi-word units but also peculiarities of the category in question.

2. Phraseology: theoretical framework

The aim of the following chapter is to present the phraseological theoretical framework in which the aims and research questions of my study are formulated. The first sections (2.1., 2.2., 2.3.) describe the discipline in both general terms and more specific terms and define the area of study. The following section (2.4.) focuses on the formal (structural) classification of phrasemes. The last section (2.5.) presents several approaches to idiomaticity (or anomaly) adopted in the fields of phraseology and lexical semantics.

2.1. Phraseology as a discipline

Phraseology as a linguistic discipline is usually said to have emerged in the 1940s in Eastern Europe and the former Soviet Union. However, it was not until the 1980s that the field attracted interest as a discipline on its own also among scholars from Western Europe and the USA (cf. Cowie, 1998: 1) and gradually developed into a recognized independent area of study. It is therefore not surprising that, for example, *The Oxford Companion to the English Language* (1992: 776) describes phraseology purely practically as "[a] way of expressing oneself; the way in which words and phrases are used, especially by particular individuals or groups." The discipline is traditionally presented as "the study of the structure, meaning and use of word combinations" (Cowie, 1994: 3168). However, it suffers from a lack of unified scope of study and terminology which can be seen, for instance, in Cowie's (1998: 7) summarizing table of terminology used for word-like units by different authors:

Author	General category	Opaque, invariable unit	Partially motivated unit	Phraseologically bound unit
Vinogradov (1947)	Phraseological unit	Phraseological fusion	Phraseological unity	Phraseological combination
Amosova (1963)	Phraseological unit	Idiom	Idiom (not differentiated)	Phraseme, or Phraseoloid
Cowie (1981)	Composite	Pure idiom	Figurative idiom	Restricted collocation
Mel'čuk (1988*b*)	Semantic phraseme	Idiom	Idiom (not differentiated)[a]	Collocation
Gläser (1988*a*)	Nomination	Idiom	Idiom (not differentiated)	Restricted collocation
Howarth (1996)	Composite unit	Pure idiom	Figurative idiom	Restricted collocation

[a] Melčuk also recognizes a so-called quasi-idiom, in which the meaning of the whole is derived compositionally from those of the parts, but where there is also 'an unpredictable addition'. An example is *bacon and eggs* (where both ingredients are often fried and the whole is traditionally served -- in the UK at least -- as a breakfast dish).

Table 1: Subcategories of word-like combinations ("nominations") as presented in Cowie (1998)

What is, however, more important, the different approaches to phraseology are reflected also in the description of the overall scope of the discipline. This is discussed in more detail in the following section.

2.2. Approaches to phraseology

There have emerged various theoretical approaches to phraseology during the recent decades. However, it can be said that two approaches have had greater impact on the present state of the discipline than others: the phraseological and the distributional (or frequency-based) perspective. The phraseological approach is represented by scholars from the former Soviet Union and other countries of Eastern Europe (Vinogradov, Amosova, cf. Granger & Paquot, 2008) and in the UK their direct follower and successor is particularly Cowie with his continuum which describes word combinations on a scale ranging from free combinations, through restricted collocations and figurative idioms to pure idioms. According to Granger & Paquot (2008: 28), one of the main objectives of this line of research is "to find linguistic criteria for distinguishing one type of phraseological unit from another and especially for distinguishing the most variable and transparent multi-word units from free combinations." Among the most common criteria in this respect belongs non-compositionality of meaning.

The second approach is represented especially by John Sinclair. In contrast with the phraseological approach, Sinclair (1987) put forward a corpus-driven method for automated identification of lexical co-occurrences. Thus, the phraseological units are not identified on the basis of a set of criteria, but on the basis of their (frequent) co-occurrence in corpora. This approach is called a distributional or frequency-based approach (cf. Granger & Paquot, 2008). In other words, while the first approach focuses on the semantic and formal properties of phraseological units and their categorization, the second draws on their syntagmatic characteristics.

The different methods of the two approaches are reflected in the assumed scope of phraseology with respect to the neighbouring disciplines of semantics, morphology, syntax and discourse analysis. Accordingly, the sphere of phraseology is described rather variably in

the works of different authors and its borders with neighbouring disciplines are fuzzy (cf. 2.3.).

One of the main differences between the phraseological approach and the distributional one appears in what they understand to be one of the key aspects of the phraseological unit, i.e. semantics. The phraseological approach works intensively with the concept of compositionality (and non-compositionality) of the meaning, whereas (non-)compositonality is of no immediate relevance for the distributional approach as it cannot use internal semantic structure as a criterion for extraction from corpora. This is due to the method of retrieving phraseological units by co-occurrence in corpora. Nevertheless, semantics plays its role in this approach as well, especially in the Firthian sense in which the meaning of a word can be defined by the words it combines with. The distributional approach thus operates with terms such as semantic restriction and preference and semantic prosody.

Although the scope of study and the method seem very different in the two approaches, Cowie (2006: 580) notes that their research lines have gradually approached each other: "[t]hese positions are not as firmly entrenched [now] as they were then. […] present-day phraseologists from quite diverse backgrounds acknowledge the benefits that can accrue from an approach which combines the advantages of access to large-scale corpus data and the value of recognizing, as part of the analytical process, the grammatical and pragmatic functions that are served by multiword units" (cf. 2.5.4.).

Apart from these two approaches, there are also other theoretical positions and it is convenient to mention the relation of phraseology and cognitive linguistics. As Gries (2008: 13) points out, "[c]ognitive grammar does away with a strict separation between lexicon and grammar. The only kinds of element the linguistic system is said to contain are symbolic units." Gries (ibid.) also emphasises that "unit status correlates positively with a speaker/hearer not analysing the internal structure of a unit." In addition to single words, this definition may easily include multi-word expressions of all kinds and even grammatical structures with no lexical specification. The cognitive approach emphasises the correlation between the frequency of occurrences and the status of symbolic units, assuming that when something is sufficiently frequent, it is stored in the brain as one unit, which is also the case with units of

phraseology. If we relate the above description with the topic of the present study, it is evident that analysing single-word lexemes in terms of phraseology is unproblematic for cognitive linguists. For the cognitive approach to phraseology, see also 3.3.1.

2.3. Phraseology and other disciplines

The structuralist approach to linguistics, from which the traditional phraseological approach emerged, operates with a system of levels of language analysis. Since the present study attempts to argue for the widening of the scope of phraseology from the syntactic level to the morphological level, the following sections summarize where phraseology borders on other linguistic disciplines.

2.3.1. Phraseology and semantics

The fuzzy area between phraseology and semantics is represented above all by the concepts of compositionality and non-compositionality (cf. 2.5.1.). In fact, semantics and the traditional phraseology overlap to a great extent in this respect, especially because semantic anomaly as one of the features of phraseological units is probably the most significant kind of anomaly (apart from syntactic frozenness and collocational anomaly, cf. 2.6.). In my opinion, this large overlap is one of the main reasons for the striking discrepancy between the practical analysis of complex single-word lexemes in phraseological terms (discussing their idiomaticity) and their absence in the "official" classifications of phraseological units (cf. Chapters 3 and 4): complex units at both levels of language constitute the same kind of material for a lexical semanticist (cf. Cruse's (2000: 69) discussion of compositionality where negational descriptors are exemplified by both *ex-lover* and *former President*), whereas phraseology is traditionally confined to word combinations.

2.3.2. Phraseology and morphology

The borderline between phraseology and morphology is of particular interest for the present study. The above definition of phraseology by Cowie (1994) excludes single-word units from the phraseological study. However, even with this clear criterion, the borderline with

morphology is fuzzy. Granger & Paquot (2008: 32-33) point out that the definition of phraseology as the study of word combinations implying that phraseological units are polylexical is problematized by the ambiguous concept of word. The word may be defined as an uninterrupted sequence of graphemes separated by spaces at either side (orthographic word) or as a "unit characterized by internal stability and uninterruptability" (Lyons, 1968: 202). Expressions such as *of course* or *letter box* may thus be included or excluded from phraseology, depending on which definition we choose. The application of the orthographic criterion is moreover complicated by the fact that one and the same compound may be spelled in three different ways, as a solid, hyphenated or open compound (*piggybank, piggy-bank, piggy bank*), and so simultaneously qualify for and fail the status of a phraseological unit. As a consequence, this inevitably leads to inconsistencies and Granger & Paquot conclude that

> One regularly has to scan through the examples given by the authors to find out whether or not (solid, hyphenated and/or open) compounds are included in the range of phraseological units covered. The traditional view either excludes compounds from phraseology altogether (Barkema, 1996: 133) or only keeps units that meet some well-defined criteria (stress, meaning, etc.). Others seem to exclude compounds written as one word, viz. solid compounds, but include open and hyphenated compounds (e.g. Mel'čuk, 1995; Gläser, 1998). In the distributional approach, all sequences made up of two or more graphic words are extracted if they meet some recurrence or co-occurrence threshold. As a result, a wide range of phraseological units are extracted, including open compounds (and possibly hyphenated ones) but excluding solid compounds.
>
> (Granger & Paquot, 2008: 61)

Similarly, Čermák (2007a: 14) illustrates by Czech examples such as *načase* (as in *je načase* "it's time") that the rules of orthography may influence the perception of expressions as either single-word units or multi-word units (since spelling *na čase* is possible as well).

2.3.3. Phraseology and syntax

As far as the borderline between phraseology and syntax is concerned, Granger (2005) notes that the fuzzy area includes collocations, especially what Benson et al. (1986) call grammatical collocations. Grammatical collocations combine a lexical and a grammatical word, such as *aim at, afraid of*. Traditionally, prepositional verbs are discussed in terms of valency, i.e. as a syntactic phenomenon (cf. Allerton, 1982).

Another fuzzy area straddling phraseology and syntax is that of compounds, which were already mentioned in the previous section in connection with morphology. In the traditional phraseological approach, the phraseological status of compounds is often determined with the help of the notion of syntactic flexibility: among the criteria used in the identification of phraseological units is the degree to which the given expression can undergo syntactic variation. According to Čermák (2007a: 82), one of the features typical of phrasemes is the "impossibility of paradigmatic substitution". This means that by "commutation test" we try to substitute one component of the possible phraseme by another component of the same or similar function.

The indistinct dividing line between phrases and compounds has been repeatedly explored by Heinz Giegerich (2004), who discusses the criteria (such as *forestress* and *attribute- or complement-head structure*) for assigning the N+N type of compound either to the syntax as phrases (e.g. the type *steel bridge*) or to the derivational morphology as having the status of compounds (e.g. the type *watch-maker*) or the possibility of whether both these types are produced in the syntax or alternatively in the lexicon. He believes that both the syntax and the lexicon are potential sources of N+N compounds but that, regardless of borderline cases, the distinction between phrases and compounds is in principle possible. However, he admits (p. 7) that "the divide between the syntax and the lexicon must be expected to be blurred, and that it must therefore be modelled as such, in order to facilitate the movement of constructions from the former into the latter through time ('lexicalization')". In this he differs from Bauer (1998: 1) who "finds that the various criteria invoked by others to motivate a syntax–lexicon split for NNs fail to correlate with each other; and he concludes that there is therefore no evidence to support any assumption of different grammatical modules being involved in the generation of NNs." Anyway, the phrasal status of some NNs (or the impossibility to exclude them from the syntax while being listed or lexicalized) opens the gate for subsuming them under phraseology.

As we can see, compounds may in the views of different authors belong to morphology, syntax or phraseology. These striking discrepancies in their classification only reflect their transitional position caused by the typological status of English.

2.3.4. Phraseology and discourse analysis

Phraseology has always had a strong link to the study of discourse, as both disciplines are concerned with multi-word units. Traditionally, phraseology put focus on language in interaction, the main domain of study being formulae (cf. Cowie 1988). With the rise of corpus linguistics, however, there came an interest in the structures of the written language. This is caused by the nature of corpus material and the shift in the scope of phraseology. The contact area between phraseology and discourse analysis is, however, not a topic of the present work, and we don't expect to find a significant link between lexical idioms and discourse analysis.

2.4. The phraseological approach to the classification of units

In order to clarify the position of lexical idioms among other types of phrasemes it will be useful to have a look at some of the standard classifications within the phraseological tradition. The present section summarizes the classification of phrasemes as put forward by several influential authors (Cowie, Howarth, Burger, Čermák) and describes an attempt to reconcile the phraseological and the distributional ways of classifying phraseological units (Granger & Paquot, 2008). As will be seen, most of these theories pay attention to referential, semantic/pragmatic and functional/communicative aspects and take the multi-word status of phrasemes for granted. Still, some of the classificatory categories may be applied even to single-word lexical idioms. I will focus on the formal features of phrasemes, i.e. their (apparent) syntactic structure. The issue of non-compositionality – idiomaticity will be mentioned only briefly and dealt with in detail in 2.5.

2.4.1. The British tradition: A.P. Cowie

Cowie's much-quoted classification (1988) divides word combinations into composites, i.e. "word combinations more or less invariable in form and more or less unitary in meaning, which function as constituents of sentences" (p. 134) and formulae, i.e. word combinations referring to sentence-like units. Composites further include the categories of restricted collocations, i.e. combinations of words which are slightly semantically anomalous, figurative

idioms, i.e. idioms based on metaphor, but "still partially analysable" (p. 135) and pure idioms, defined by their non-compositionality, where "the literal senses […] do not survive alongside their figurative ones." (p. 135). Later Cowie (2001) divides formulae into routine formulae (conveying a communicative function) and speech formulae (organizing the utterance and indicating speaker's or writer's attitude). Figure 1 below summarizes Cowie's hierarchy of word combinations:

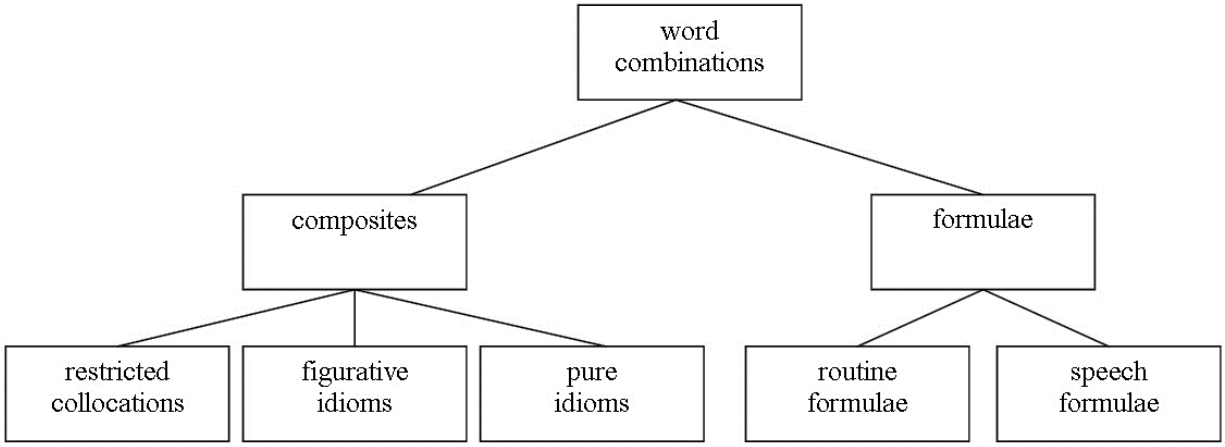

Figure 1: Cowie's (2001) classification of phraseological units

As can be seen in Figure 1, Cowie restricts phraseology to word combinations, excluding thus other types of combinations from phraseology. The second level in the hierarchy branches word combinations depending on their form (below vs. on/above the sentence level), with the lowest level different for each group: composites are classified according to their semantic compositionality, whereas formulae are classified according to their function (communicative vs. textual).

2.4.2. The British tradition: Peter Howarth

Peter Howarth (1998) distinguishes functional expressions and composite units, which is analogical to Cowie's hierarchy. Composite units are described by him (p. 27) as units having "a syntactic function in the clause or sentence and are generally best seen as realizations of phrase structures such as prepositional phrases, noun phrases, etc." Composite units are further divided into lexical composites, consisting of two open-class units (e.g. V+N, *make a*

claim or A+N, *ulterior motive*), and grammatical composites, consisting of an open class unit and a closed class unit (e.g. Prep+N, *in advance*). Functional expressions, on the other hand, have a distinct role in discourse (e.g. complete utterances such as proverbs, or discourse markers such as conjuncts). Figure 2 below summarizes Howarth's hierarchy of word combinations:

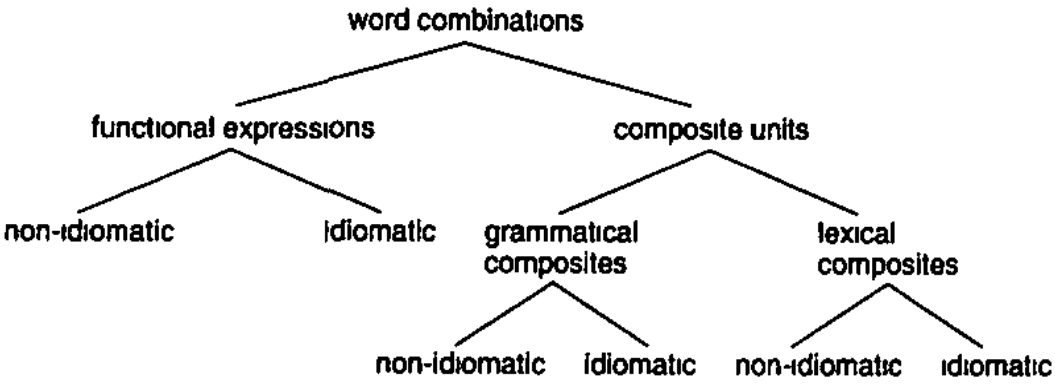

Figure 2: Howarth's phraseological categories (1998)

Figure 2 also shows that all three types of word combinations are then classified according to the semantic compositionality as non-idiomatic or idiomatic (see 2.5.4. for detail).

In summary, although hierarchies presented by Cowie and Howarth are not identical, they classify word combinations using the same criteria: function (communicative, textual), form (phrase, sentence) and meaning (non-idiomatic, idiomatic). Moreover, the two classifications restrict phraseology to word combinations and do not mention phraseological aspects of single-word units at all. The same is true of Burger's classification, which is representative of one influential strain of Continental thought.

2.4.3. The Continental approach: Harald Burger

Burger's (1998) concept of phraseology can be taken as another example of a structuralist approach. Like Cowie and Howarth, Burger limits the scope of phraseology to multi-word units when he posits two basic properties of phrasemes: polylexicality (Polylexikalität) and fixedness (Festigkeit), accompanied by idiomaticity (Idiomatizität), which is however not as

crucial as the first two properties (pp. 14-15). Burger's classification focuses on the function of the unit in discourse. He distinguishes three main groups: referential phrasemes, which refer to objects, processes or facts, and are further divided into nominative (i.e. expressions on the level of clause elements) and propositional phrasemes (i.e. expressions on the level of sentence). This part of the classification is similar to that of Cowie (1988) and other structural linguists. Burger then classifies phrasemes into collocations, partial idioms and idioms, again similarly to other traditional classifications (cf. Table 1 in 2.1). However, in addition to the class of referential phrasemes, Burger also distinguishes structural phrasemes whose function is purely grammatical and which correspond roughly to multi-word prepositions and linking adverbials (his examples include *in Bezug auf* "in relation to" *sowohl - als auch* "as well … as …")[1] and communicative phrasemes which fulfil an interactional function and are mostly used in conversation (*Guten Morgen* "good morning", *ich meine* "I mean"). They correspond roughly to Cowie's routine formulae. Figure 3 presents the basic classification:

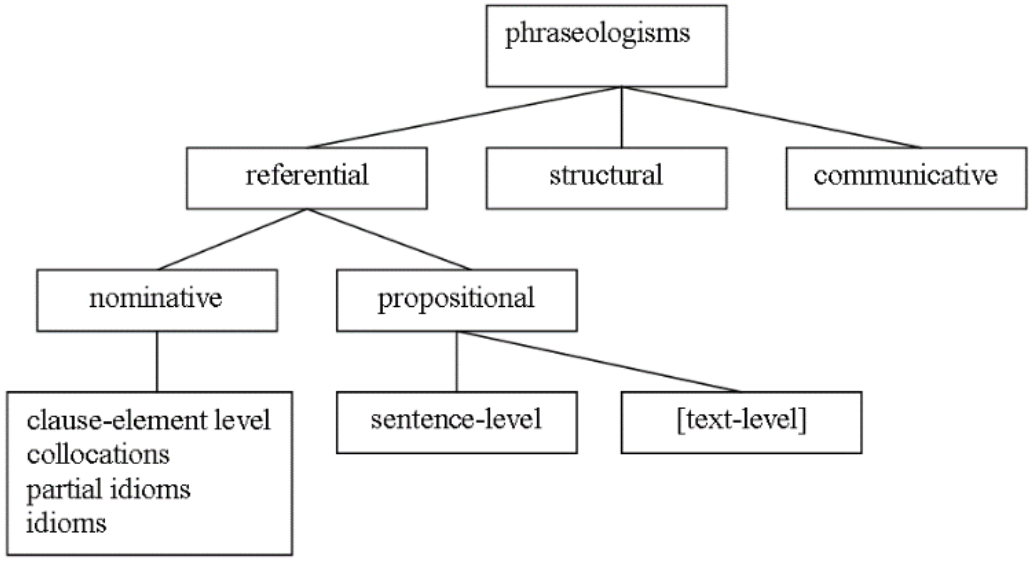

Figure 3: Classification of units of phraseology according to Burger (1998)

Figure 3 illustrates that, although the hierarchy is similar to Cowie's (1988) hierarchy, word combinations are in the first step classified according to their function and not the form.

[1] Note that although they have a similar function, Burger's structural phrasemes are different from Cowie's speech formulae: Structural phrasemes are below the sentence level, whereas speech formulae are on the sentence level.

In a later text Burger et al. (2007) attempt to unify terminological differences in the discipline and propose terms which should be used. The term suggested for the whole discipline is *Phraseology*, which should also include *Paremiology*, i.e. the study of proverbs, which is sometimes treated separately. Burger et al. also mention an interesting phenomenon linked mostly to the development of the distributional approach:

> In the beginning phase of phraseology research, there was a tendency to circumscribe the research field for purposes of consolidation. However, recently, the opposite has been the case, especially because questions that are closely related to conventional phraseological questions, but not identical, have been raised by corpus linguistics. Data on a scale that was unheard of until now, new types of data and new forms of analysis lead to new questions or to restatements of old questions.
>
> (Burger et al., 2007: 11)

The tendency to broadening the scope of phraseology is also relevant for the present topic of lexical idioms. As far as the basic unit of phraseology is concerned, Burger mentions several terms used quite generally: *phraseologism, phraseme* and *set phrase*. He concludes that *phraseme* is the most convenient term internationally, but that *set phrase* can be expected to be used as well because of its broad use in English. However, Burger et al. (2007: 12) also admit that "[t]he term 'phraseme' has the disadvantage that its suffix -eme emphasizes the systematic aspect (compare "phoneme", "morpheme", "lexeme", "texteme"). It is probably impossible to cover all formulaic aspects other than idioms, collocations, etc. with this term. The term "phraseme" can only be used with restrictions as soon as the narrow subject area of phraseology is left behind." Therefore, the term "phraseme" is, according to Burger et al., suitable for the basic unit of phraseology in the narrow sense (including idioms and collocations), whereas another term, for example *formulaic language*, can be used in wider conceptions. The proposed subcategories of the phraseme (p. 15) are *idioms* ("semantically marked set phrases"), *collocations* (units "with a weak or non-existent semantic reinterpretation"), and proverbial expressions.

2.4.4. An alternative Continental approach: František Čermák

Čermák's original classification of phraseological units was developed for Czech for the purposes of a pioneering project, the compilation of *Slovník české frazeologie a idiomatiky* I-IV (2009), of which Čermák was the main editor. Although designed for Czech, his classification based on systematic and consistent criteria is sufficiently general to be in principle applicable and relevant to English as well. As it is, moreover, one of the starting points for this study, I include it alongside classifications focusing on English. In contrast to the three theorical stances mentioned above, Čermák (2007a) focuses essentially on idiomatic structures only; non-idiomatic structures are not included in the description. However, the scope of units described within phraseology by Čermák is roughly the same as with Cowie, Howarth and Burger. This is because Čermák defines idioms by reference to a combination of anomalous properties, including not only semantic aspects, but also formal anomalies and anomalies in collocability (cf. 2.6.). He calls the basic unit of phraseology *phraseme* when discussing the unit from the formal point of view and uses the term *idiom* when he refers to the semantic features of the unit. If a unit combines both formal and semantic anomaly, it can be referred to by both terms. For simplification, the combination of both terms, phraseme / idiom, is often used in his texts (2007a).

Phrasemes / idioms (PIs) are classified according to the language level they belong to as lexical PIs, collocational PIs and propositional PIs. Combinatorial possibilities of components (Čermák, 2017) of PIs are outlined in Table 2:

level of components level of PI		A morphematic	B lexical	C collocational	D propositional
1	morphematic				
2	lexical	+			
3	collocational	+	+		
4	propositional	+	+	+	
5	polypropositional single-subject			+	+
6	polypropositional intersubject				+

Table 2: Combinatorial possibilities of phraseological components

Lexical PIs (referred to as *lexical idioms* in my study) are single-word units with anomalous combination of morphemes (or stems, bases). They will be dealt with separately in Chapter 4. Collocational PIs are multi-word expressions functioning below the sentence level. They are further divided into grammatical PIs (containing functional words; the whole expression is a member of a closed class - prepositions, conjunctions, particles, pronouns), nominal PIs (which function as nouns), modification PIs (which function as adjectives) and verbal PIs (which function as a whole predicate). Propositional PIs are then divided into propositional (consisting of one clause) and polypropositional (consisting of more than one clause), single-subject propositional PIs contain one speaker, whereas intersubject PIs are represented by a dialogue. In addition to these groups based on the word-class function of the expression, Čermák (2007a) also describes two subtypes of PIs distinguished by their structure which can be both verbal and non-verbal: similes and binomials. The formal aspects of Čermák's hierarchy are illustrated by Figure 4:

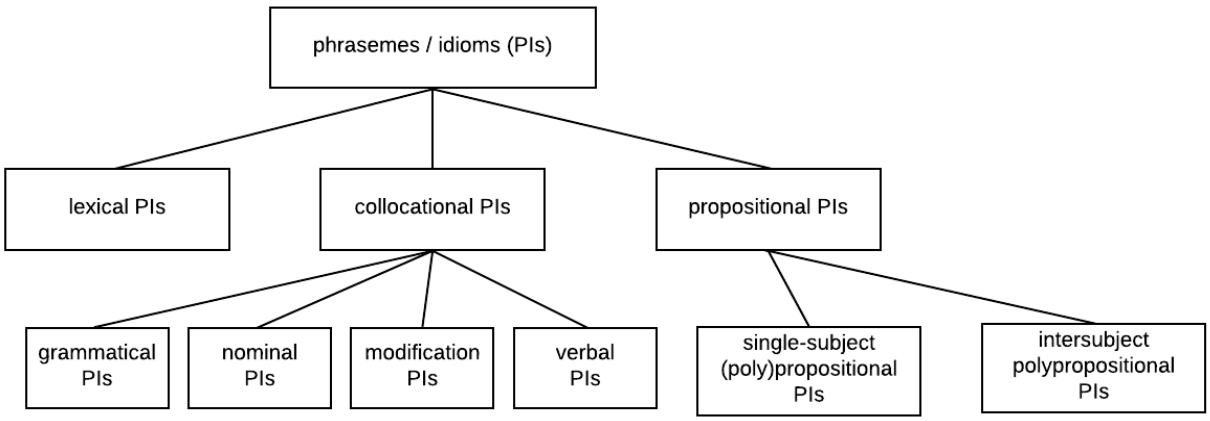

Figure 4: Čermák's classification of phrasemes / idioms

2.4.5. Attempting a classificatory synthesis: Granger and Paquot

After reviewing the two approaches to phraseology, Sylviane Granger and Magali Paquot (2008) propose a reconciliation of the traditional classifications and the changes brought about by the distributional approach. They suggest "making a clear distinction between the

two typologies: keeping one for the automated extraction and one for linguistic analysis" (p. 41). This means that terms such as *n-grams* (i.e. continuous sequences of 2 or more words) and *co-occurrences* (i.e. discontinuous combinations of two words) are to be used for the extraction, whereas the term *collocation* should keep its traditional meaning. Figure 5 shows their classification of phrasemes, which is partly an extension of the classification presented by Burger (see above):

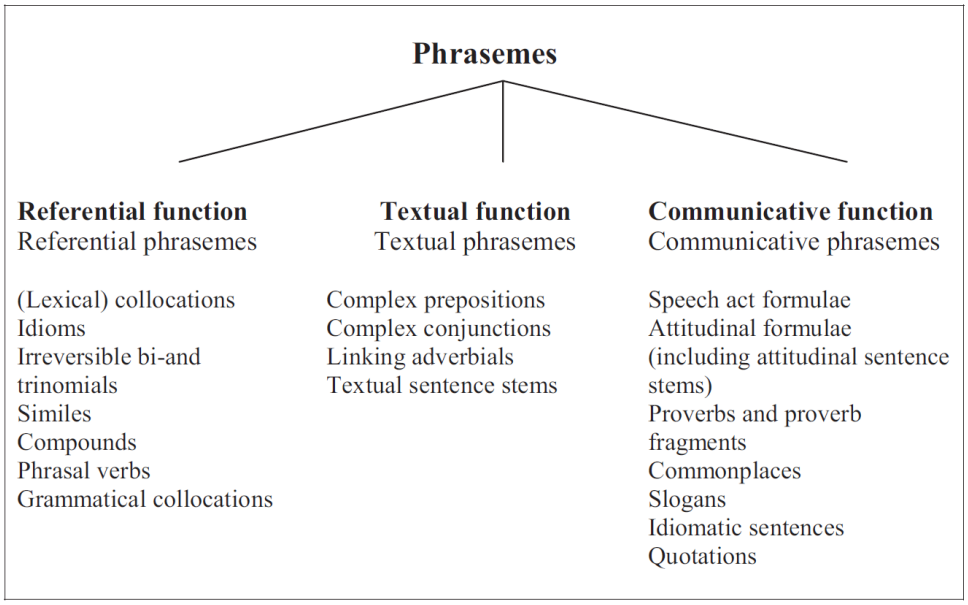

Figure 5: Classification of phrasemes (Granger & Paquot, 2008)

They retain the basic three categories introduced by Burger (1998), textual phrasemes being extended by his 'structural phrasemes'. Of the three, referential phrasemes are the category of the greatest importance for the present topic. They "are used to convey content message" (p. 42). They include a subset of categories represented in Figure 5. What is different from the hierarchies discussed above is that idioms are defined partially on the basis of their form as "restricted to phrasemes that are constructed around a verbal nucleus" and partially on the

basis of their (non-compositional) meaning (p. 43). In addition to collocations and idioms included in common with some of the other hierarchies as well, there are irreversible bi- and trinomials, which are defined as "fixed sequences of two or three-word forms that belong to the same part-of-speech category and are linked by the conjunction 'and' or 'or'", and similes defined as "stereotyped comparisons". Another interesting point is that the authors explicitly include compounds. In this respect, the authors cross the borders between word combinations and morpheme combinations, acknowledging to a certain extent the existence of lexical idioms. The category of textual phrasemes describes means used for structuring and organizing the text, which is in accordance with Burger's classification. Finally, communicative phrasemes (p. 42) are "used to describe feelings or beliefs towards a propositional content or to explicitly address interlocutors, either to focus their attention, include them as discourse participants or influence them."

2.5. Idiomaticity

After examining the structural aspects of phraseological units in Section 2.4., we may now proceed to the crucial feature of phrasemes and look into the problem of idiomaticity. Although Čermák (2007a, 2007b) uses the term *idiomatics* (and similarly Kavka (2009) employs the term *idiomatology*) as the name for the discipline alongside *phraseology* in order to emphasize the semantic side of the units studied rather than their phrasal structure, I have decided to avoid the term and speak only of *idiomaticity* as the primary defining feature of phraseological units. This is partly because the term *idiomatics* is not very well-known and so rarely used, but above all because the issue is equally relevant outside the strictly phraseological context and is often treated by both phraseologists and semanticists.

The literature mentions a number of features that characterize idiomaticity, making it possible to differentiate idioms from distributional units such as lexical bundles (*at the end of*) or collocations (*world view*). I will focus especially on those which are relevant for this study, (non-)compositionality and anomaly, but also briefly refer to several others.

2.5.1. Compositionality and non-compositionality

Non-compositionality is one of the most common features mentioned in connection with phraseological units. However, to define non-compositionality entails defining compositionality first.

2.5.2. Compositionality in terms of lexical semantics

Alan Cruse (2000: 67) defines the principle of compositionality as follows: "The meaning of a grammatically complex form is a compositional function of the meanings of its grammatical constituents." There are three claims which, according to Cruse, are incorporated in this definition:

(i) The meaning of a complex expression is completely **determined** by the meanings of its constituents.

(ii) The meaning of a complex expression is completely **predictable** by general rules from the meaning of its constituents

(iii) Every grammatical constituent has a meaning which contributes to the meaning of the whole.

At the same time, Cruse (p. 70) stresses that "[t]he principle of compositionality as set above is not universally valid" as there exist expressions where not all grammatical constituents represent an identifiable part of the meaning of the whole and reformulates the principle of compositionality stated above, introducing the notion of semantic constituent: "The meaning of a complex expression is a compositional function of the meanings of its semantic constituents, that is, those constituents which exhaustively partition the complex, and whose meanings, when appropriately compounded, yield the (full) global meaning.". Cruse also proposes a *contrast test* which helps to identify semantic constituents. According to this test, semantic constituents "can be substituted by something else (belonging to the same grammatical class), giving a different meaning." (Čermák (2007: 78) uses the term *collocational paradigm* for the set of possible constituents filling a slot in an expression.) The substitution test includes a second step to show that the meaning is an inherent part of the constituent (Cruse, 2000: 71): "[a]t least some of the contrasts of the meaning produced by

substitution in one context should be reproducible using the same items in a (formally) different context." Cruse (p. 72) illustrates the principle by a phrase representing a free combination of words, *the cat sat on a mat*, but also on the level of word, contrasting by the test between compositional *disapprove* and non-compositional *disappoint* (where "adding *dis-* does not create an opposite as it does with *approve* and *mount"* and several other examples (*recount* "count again" or "narrate", *report, receive, revolve*, *blackbird* and *strawberry*). We can add our own example on the level of word to illustrate the test in full: *merciful* consists of two grammatical constituents, *mercy* and *-ful*. *Mercy* is a semantic constituent because it can be substituted by something else in the given syntagma: e.g. *hate* in *hateful*. A similar contrast can be reproduced in a different context: *He showed mercy* vs. *he showed hate*. The same test can be applied to *-ful:* It can be substituted by *-less: merciless* and the contrast between *-ful* and *-less* is kept in e.g. *joyful – joyless*. The word *merciful* can thus be, according to Cruse's criteria, considered fully compositional.

Cruse (ibid.) defines an idiom on the basis of the criteria stated above as "a type of grammatically complex expression not all of whose grammatical constituents are semantic constituents." In his description he focuses on multi-word units (e.g. *to pull (someone's) leg*, *to paint the town red*) and identifies five properties of elements of idiomatic expressions which are not functioning as semantic constituents:

1. Elements are not separately modifiable without loss of idiomatic meaning (**She pulled her brother's <u>left</u> leg.*)
2. Elements do not co-ordinate with genuine semantic constituents (**She pulled <u>and twisted</u> her brother's leg*)
3. Elements cannot take contrastive stress, or be the focus of topicalizing transformations, and the like (**It was her brother's LEG that she pulled. *What she did to her brother's leg was pull it.*)
4. Elements cannot be referred back to anaphorically (**Mary pulled her brother's leg; John pulled it, too.*)
5. An idiom does not survive the substitution of any of its constituent elements by a synonym or near synonym (**The poor old chap kicked the pail.*)

In addition, Cruses stresses that some aspects of grammar associated with idioms may not be part of the idiom, which means that changing these aspects does not destroy the idiom (e.g. passive in *his leg was being pulled continually by other boys*). However, this rule is not valid generally for all idioms and the same aspect may be elsewhere part of the idiom which cannot be changed (**The bucket was kicked by him*).

It seems that these principles are not very useful for the identification of lexical idioms. If we try to apply the rules to two meanings of a word, one compositional and one idiomatic (judged by intuition), *revisit* "to visit (someone) again" and *revisit* "to consider, inspect (e.g. a topic) again", we can see that the only rule which is partly applicable to the non-idiomatic use is the rule number three, as semantic constituents of words can take contrastive stress:

A: *I revisited Paris last summer.*

B: *Was it your first time in Paris?*

A: *No, I said I* RE*visited Paris.*

Re- in the second sense of *revisit* cannot be stressed in this way (**We have to* RE*visit the issue.*). However, it is questionable whether putting a contrasting stress on the second semantic element, *visit* (which is a base morpheme), would sound natural in any context. The reason why these rules are not applicable is that fixedness of elements within a word, i.e. the degree of lexicalization which the test shows, is high for all units on the level of words. The only blurred line which touches out present topic is the line between the level of words and the level of phrases (e.g. *black bird* vs. *blackbird*).

In addition to idioms, Cruse distinguishes three other types of expressions: frozen metaphors, collocations and clichés. Collocations and clichés are not substantial for the topic of lexical idioms and therefore only the third type will be discussed here. Frozen metaphors do not pass the contrast test described above but show only some of the properties of idioms. In particular, a constituent of a frozen metaphor may be substituted by a synonym without a complete loss of non-compositional meaning (e.g. *The ball's in your court now. > The ball's on your side of the net.*). Cruse (p. 75) clarifies the status of these expressions pointing out that "the literal meanings of the constituents of idioms are not always inactive or irrelevant to the idiomatic meaning. The degree of relatedness between literal and non/literal meanings of idioms varies

continuously from none to all to such a high degree that the expression falls into a shadowy border area between idiomaticity and full compositionality."

This "shadowy border area" is one of the most problematic part for the present study, especially if we take into account the fact that it is not possible to use Cruse's five properties of (multi-word) idioms as the test of idiomaticity when searching for lexical idioms.

Cruse (pp. 77-80) goes on to present other problematic areas linked to compositionality. The first problem concerns expressions consisting of a descriptive adjective and a noun. In this case, the locus of the quality is not the same in all cases, cf. *a red apple* "an apple red on the outside" and a yellow peach "a peach yellow on the inside". However, although these *active zones* are not predictable by a simple rule and must be learned (the hearer usually needs some knowledge of extralinguistic reality), these expressions are intuitively considered non-idiomatic and they also pass the recurrent contrast test described above (*a red apple* vs. *a green apple*)

The second problem related to complex categories (categories consisting of two or more simple categories merged together) may be relevant for the present topic as well. In summary, the prototypical example of the complex category is neither the prototypical example of the first, not of the second simple category, but it is rather an example which manifests the greatest proportion possible of the first category and at the same time of the second category. This is exemplified by the complex category PET FISH. The respondent often name *guppy* as the prototypical PET FISH although it is neither the prototypical *pet* nor *fish*. Instead *pet fish* are described as a "those fish nearest to the prototype pets" (Cruse, 2000:79). This phenomenon described in terms of the prototypical theory must be taken into account during the identification of lexical idioms of all kinds, especially idioms consisting of two lexical stems (compounds).

2.5.3. The Continental phraseological approach to (non-)compositionality

Harald Burger's (1998: 31-32) account of idiomaticity can be presented here as the first example of the phraseological approach to (non-)compositionality. He defines idiomaticity as a discrepancy between the phraseological meaning and the literal meaning of the parts the

phraseme is composed of. In addition to his basic definition, he notes that idiomaticity is a gradual property of idioms, which means that the discrepancy is higher for some idioms and lower for others (cf. Cowie, 2.4.1.).

In his classification, Burger (p. 56) proposes a system of three grades: idioms, partial idioms and non-idioms, although he emphasizes that the transition between these categories is gradual. Another problem which is described by Burger and which must be taken into account in the present study, is that the so-called *free meaning* (freie Bedeutung, i.e. non-idiomatic meaning) may be sometimes difficult to describe because a word has very often several related meanings or there may be homonymous words with different unrelated meanings. It is therefore important to distinguish between expressions which are idiomatic because their lexical meaning is different from the meaning of their parts and expression which are themselves non-idiomatic but consist of parts which are already used in an idiomatic sense. In addition, Burger describes phrasemes which may be read both literary and idiomatically with little difference in meaning (e.g. *to shrug one's shoulders*).

2.5.4. Compositionality in the British phraseological approach to idiomaticity

Peter Howarth (1998: 28) approaches idiomaticity similarly to Cowie and Burger, introducing a continuum with free combinations on one side and pure idioms on the other. He identifies four distinct classes: free combinations, restricted collocations, figurative idioms and pure idioms:

	free combinations	restricted collocations	figurative idioms	pure idioms
lexical composites verb + noun	*blow a trumpet*	*blow a fuse*	*blow your own trumpet*	*blow the gaff*
grammatical composites preposition + noun	*under the table*	*under attack*	*under the microscope*	*under the weather*

Figure 6: Howarth's collocational continuum

Free combinations "consist of elements used in their literal senses and freely substitutable". The definition corresponds to Cruse's claims about prototypical compositionality and the test

of semantic constituents, adding the idea of "literal sense" which is not mentioned by Cruse. Restricted collocations "have one component […] that is used in a specialized, often figurative sense only found in the context of a limited number of collocates." (ibid.) Figurative idioms are described as having "metaphorical meanings in terms of the whole and have a current literal interpretation" (ibid.). Finally, pure idioms display a "unitary meaning that cannot be derived from the meanings of the components and are the most opaque and fixed category." (ibid.). The examples of these categories are provided in Figure 6.

As Granger (2005: 1) points out in her summary of the phraseological approach, "[o]ne of the main preoccupations of linguists working within this tradition has been to find linguistic criteria to distinguish one type of phraseological unit from another (e.g. collocations vs. idioms or full idioms vs. semi-idioms) and especially to distinguish the most variable and transparent multi-word units from free combinations, which only have syntactic and semantic restrictions and are therefore considered as falling outside the realm of phraseology (Cowie 1998: 6)."

2.5.5. The compromise distributional perspective on idiomaticity

Rosamund Moon (2015) presents a view on phraseology and idiomaticity which combines achievements of both phraseological and distributional approach. As Moon notes (p. 121), she adopts "a middle position" with respect to idioms, describing them as multi-word items "which are problematic because of their semantics: potentially ambiguous, often figurative, and also often evaluative and connotative." Three basic criteria for identification of multi-word items and idioms are according to Moon (p. 122) *institutionalization*, *fixedness* and *non-compositionality*. Institutionalization is "the extent to which a string of words recurs", and it can be measured by corpus data (the frequency of the string). However, Moon also adds that an institutionalized string of words should also be recognized "as a holistic sequence in the lexicon in order to exclude very frequent freely-formed strings such as *in the middle of* or *it is possible to*."

Fixedness is either paradigmatic, where addition, omission or substitution of components is impossible, or syntagmatic, referring to restrictions on sequencing and/or regular grammatical operations.

Non-compositionality is understood traditionally as an instance where the unitary meaning is not derived from the meaning of the components. Moon (p. 126), as a corpus linguist, notices the fact that the range of possible semantic anomalies is wide: some figurative expressions are quite transparent (*cut corners*), some can be decoded simply by conventional connotations (her example is *smell a rat* "be suspicious about something"), other depend on both context and the interpreter. She adds: "For items like a piece of cake, the idiomatic meaning might be guessed from context, but might not; it could be interpreted as 'something pleasant or indulgent' rather than 'something easy'. Thus non-compositionality is subjective, depending on individuals' linguistic and metaphorical competence and their decoding of component words." Moon also mentions potential ambiguity of idioms (idiomatic vs. literal meaning), which is normally disambiguated by context, and genuine ambiguity of expressions such as *shake hands* or *raise one's eyebrows* (cf. Burger in 2.5.1.2.).

As far as the origins of idioms are concerned, Moon (ibid.) lists, apart from idiosyncratic idioms often of uncertain origin, also idioms based on a conceptual metaphor (cf. Lakoff & Johnson 1980, see also Benczes, in 3.3.1.) and metonymic idioms (*lend a hand*).

2.6. The concept of anomaly

Čermák (2007: 76) claims that semantic non-compositionally is not the only, or even exclusive, feature of idiomaticity and introduces the concept of multiple anomaly, distinguishing between regular and anomalous combinations in language. Regular combinations are "combinations governed by analogous rules". This class includes combinations on all levels of language description: combinations of morphemes, lexemes, collocations and sentences (cf. Table 2). The rules governing regular combinations may be both semantic, "based on semantic compatibility of the combined elements and the meaningfulness of their resultant combinations", and formal and grammatical, i.e. syntactic rules and collocational rules.

Anomalous combinations, on the other hand, are not regular in all of the described aspects. As such these combinations belong to the field of phraseology / idiomatics, which goes "a step further in the area of combinatory outcomes of the extensions of combinatory possibilities in

language, crossing the boundaries and operating in an area of combinations that, according to the standard rules of language, cannot or should not take place" (ibid.).

Anomalies can be of several kinds (pp. 81-83): paradigmatic anomaly is closely related to collocability and the meaning of the combined elements: while regular complex units permit free combination of elements which are semantically compatible, anomalous combinations display limited combinatorial possibilities and sometimes the class of possible elements contains only one member (monocollocability). Often, the meaning of elements of these anomalous combinations is also anomalous: the elements do not have the same meaning as in other environments. Collocational anomaly can be tested by the commutation test (cf. Cruse's contrast test in 2.5.2.) which examines whether the elements in the complex unit are substitutable by other elements sharing the same general function (in Čermák's terms, by other elements belonging to the same virtual paradigm). Anomaly can also affect grammatical behaviour: anomalous units sometimes do not allow grammatical transformations, such as negative, passive, change of grammatical number, etc. In other words, the units are grammatically frozen.

2.7. Criteria of idiomaticity

Čermák (2007b: 20) proposes two defining features of phrasemes and idioms: multi-component character and anomalous character of their structure (including both semantic anomaly, i.e. non-compositionality, and varying degrees of grammatical and collocational anomaly). It is this first feature which makes it possible to include single-word units consisting of more than one morpheme to be included in the study of idioms (for Čermák's account of lexical idioms see Chapter 4). The second defining feature, the anomalous character of the structure, then serves to identify which units from these areas can be regarded as idioms. Although many influential theories define idioms as based on anomalous meaning only (cf. Cruse, 2000 and Cowie, 1998), Čermák explicitly states that anomalous units may break not only semantic rules (i.e. display non-compositinality), but also collocational and formal rules of combinability.

An idea similar to Čermák's collocational and formal/grammatical anomaly is behind Gläser's (1988: 268–269) transformation tests which she applies to test whether an item is a true idiom

or not. She holds that true idioms fail all the tests, in other words their idiomatic meaning cannot withstand the transformations, while the meanings of regular phrases can survive them. She distinguishes two groups of tests, lexical transformation tests and grammatical transformation tests. The principal lexical tests are

i. augmentation (addition of lexical constituents)

ii. elimination (deletion of constituents)

iii. substitution (replacing a constituent by a semantically-related word), and

iv. permutation (rearranging constituents whose order is fixed)

The main grammatical tests are

i. blocking of predication

ii. blocking of the formation of comparative and superlative forms of adjectives,

iii. blocking of nominalization, and

iv. blocking of passivisation

The literature mentions other features or criteria of idiomaticity, some of which are closely related to non-compositionality, such as semantic transparency/opaqueness, syntactic analysability (allowing for flexibility, i.e. departure from the canonical form of an idiom), salience and adherence to truth conditions. Salience is the speaker's (subjective) belief what a lexical item means when asked about its meaning. To quote Giora (2002: 490–491), "[t]o be salient, meanings of words, phrases, or sentences (e.g. the conventional interpretations of idioms or proverbs) have to be coded in the mental lexicon and, in addition, enjoy prominence due to their conventionality, frequency, familiarity, or prototypicality. Meanings not coded in the mental lexicon (e.g. conversational implicatures constructed on the fly) are nonsalient." The adherence to truth conditions as an idiomaticity criterion is applicable in the case of expressions open to both literal and transparent metaphoric readings, e.g. *to wave a red flag before a bull*. To quote Gill (2011: 22), "The link between the phrase and its idiomatic meaning is perfectly transparent, but what makes the former meaning idiomatic (or metaphorical) rather than compositional (or literal) is that it is untruthful." Finally, related to Čermák's collocational regularity/anomaly is the concept of 'collocational harmony' (cf. Gill, 2011: 23): "A word (or

collocation or phrase) is interpreted literally when it is in *collocational harmony* with its salient context." Needless to say that the different features of idiomaticity operate together, not separately, and that they form clines or scales of idiomaticity rather than discrete categories.

2.8. Functions of idioms

Čermák (2007a: 89-91) claims that the basic function of idioms is the denominative function. There are three possible relations of idioms to regular language: firstly, idioms as such may serve "as the primary and monopoly means of expression (usually also highly economical) and regular language is not used for the given meaning". Secondly, there is "a parallel expression [of the given meaning] competing with the regular one", and thirdly, the idiom "does not obtain for the given meaning".

In addition to denominative function, closely linked to the notional part of the meaning, Čermák also emphasises the pragmatic (connotative) function of idioms. He adds that the pragmatic component of meaning is characteristic of idioms. Among the pragmatic aspects, it is especially the expressive, symbolic and evaluative components which are prominent.

2.9. Idioms vs. terms

One of the issues debated in the literature, is how multi-word terms relate to idioms with which they seem to share at least some of their features. Čermák (2007a: 220-227) discusses various types of terms and their relation to idioms. First, it is important to realize that terminology is a specific part of vocabulary with specific qualities. Čermák examines which properties terms have in common with idioms, and his findings are illustrated in Table 3:

	term	phraseme / idiom
stability	+	+
nominative power	+	+
monosemy	(+)	-
semantic transparentness	+	-
regular formation	+	-
denotatively precise	+	-

contextual non-ambiguity	+	-
membership in a discipline	(+)	(-)
pragmatic aspects	-	+
original metaforicity	(-)	+
absence of synonym	(-)	-

Table 3: Differences between terms and PI (Čermák, 2007a: 226)

As we can see from the table, there are not many instances where terms and idioms behave in the same way. However, Čermák also points out that in this respect there are differences between terms which stand in the centre (i.e. prototypical instances of terms) and terms which are at the periphery (i.e. less prototypical examples). The latter type is represented, for instance, by terms whose naming is based on metaphor, i.e. professional slang terms, folk terminology, various unofficial terms. These peripheral terms are called by Čermák quasi-idioms and they are placed in the fuzzy area between terminology and phraseology.

3. Lexical idioms as isolated phenomena

Although the idea of lexical idioms, i.e. single-word complex expressions with idiomatic meaning, runs counter to the prevailing standard definition of phraseology as the study of lexical units composed of at least two words, the fact is that in the literature not dealing with phraseology it is paradoxically not difficult to find mentions of both idiomatic (phraseological) derivatives and idiomatic compounds. This results in a peculiar situation when authors have no problem acknowledging idiomatic derivatives or compounds as isolated phenomena but there is no account of phraseology (with the exception of Čermák's) that would explicitly single out lexical idioms as a separate class of phraseological units.

For this reason, I have devoted one chapter to these mentions to idiomatic derivatives and compounds as isolated phenomena outside the phraseological literature and another chapter to lexical idioms integrated in the description of phraseology and taken account of in the classification of phraseological units. This latter approach is represented basically by just one author, F. Čermák (1982, 2007a, 2007b), and the problem is that his approach is largely tailored to Czech. The present work is therefore an attempt to apply his idea of lexical idioms as a special category of phraseological units to English and elaborate it in terms of the specific situation in English due to typological differences between Czech and English.

A preliminary search for mentions of idiomatic/phraseological derivatives and compounds in literature had showed that there are far more references to idiomatic compounds than to idiomatic derivatives. This raises the question of whether it is a mere chance or whether there are some deeper reasons for this disproportion. Čermák's own definition of a phraseological unit (cf. 2.4.4.), which essentially rests on three types of anomaly: semantic (combination of meanings incompatible in regular language, and non-compositionality of meaning), grammatical (restricted grammatical variability, formal frozenness) and lexical (restricted collocability) does not provide an immediate clue as to why there should be more idiomatic compounds than derivatives. Moreover, the last two types of anomaly, grammatical and collocational restrictions, are of limited value when it comes to complex words and the distribution of their constituents. This suggests that although in the identification of lexical

idioms both semantic and formal factors need to be taken into account, semantics is likely to be the principal indicator of idiomaticity.

In order to account for the different distribution of idiomatic derivatives and compounds in non-phraseological literature, I will start with one study dealing with word-formation which offers a possible explanation of why it should be, so using the theoretical framework of semantic coindexation, and only then will I give a necessarily brief overview of references to idiomatic derivatives and compounds made in other than phraseological contexts.

3.1. Idiomatic compounds and derivatives in terms of semantic coindexation

Rodriguez and Rio-Torto's study (2013) compares several types of derivatives and compounds in Portuguese with the aim to explore the way meaning construction occurs in derivation and compounding. It addresses three questions: how do derivatives and compounds get their meaning, which factors are involved and, most importantly for my study, does the semantics of derivative and compounds constituents follow the same rules? It also acknowledges that both derived words and compounds may have compositional and idiomatic meanings, i.e. meanings either computable or not computable from the meaning of their constituents.

The authors start form the assumption that the construction of meaning in word-formation follows from semantic coindexation independently of syntax. They posit that coindexation operates between semantic features of the constituents (affix and base or compound bases) and those of the Fillmorean 'maximal semantic frame' (schematizations of particular situation types or scenarios) which are associated with them. Next they assume that semantic coindexation between these features is governed by the degree of semantic similarity between them which, in effect, "prevents chaotic linking between features, because it only allows the linkage of those that best fit semantically with each other". In addition, they assume (following Jackendoff) that the meaning of complex words derives not only from the lexical constituents involved, but also from other sources of information, referential and /or pragmatic.

Starting with derivatives they assume that affixes are not mere formal operators of word-formation rules, but that they exhibit semantic features, which is why there are combinatory constraints between affix and base. The features are activated only when the affix combines with the base. The semantic contribution of an affix becomes evident by comparing other derivatives of that affix and derivatives of other affixes following the same rule. The differences between the meanings of derivatives are due to the coindexation of semantic features of each affix (its semantic structure) with semantic features of the base. The semantic features of the base follow from its semantic structure (e.g. 'event' in deverbal derivatives) and its lexical-conceptual structure. The authors conclude that rather than occurring at the level of argument structure, meaning construction in derivatives is based on coindexation between features of the affix, the semantic features belonging to the lexical-conceptual structure of the base, and those of the maximal semantic frame associated with it.

Likewise the meaning of a compound is related to the meaning of its constituents. Coindexation based on the features of each constituent governed by the principle of semantic plausibility, ensures the maximal compatibility between the meanings involved. The meaning is further specified and adjusted by the lexical-conceptual structure of each constituent and the relation between them and by referential and pragmatic constraints following from the 'maximal semantic frame' associated with each of the constituents, and the plausible semantic and grammatical relations. Moreover, the meaning may be subject to figurative constraints based on figurative mechanisms which supply semantic coherence "when denotational or objective tools are overlooked". These referential and/or pragmatic and figurative factors may result in idiomaticity and weak or opaque compositionality.

How does this type of interpretation of meaning construction in complex words relate to the apparently different distribution of idiomatic derivatives and idiomatic compounds? In both non-idiomatic and idiomatic derivatives and compounds the construction of the meaning derives from semantic coindexation (ruled by the compatibility principle) of the features of the constituents which is further articulated by the semantics of the constituents and their maximal semantic frame. The more straightforward the compatibility of features is the more transparent the derived word and conversely the more complex the meaning construction the higher is the idiomaticity. The set of lexical-conceptual structure (LCS) features associated

with a lexical base is potentially wider than the set of features associated with an affix and consequently the feature combinability potential of two lexical bases of a compound is even greater. Since compound constituents represent "two LCS universes that articulate and enrich each other, the meaning structure in compounding tends to be freer than in derivation" and "[by] default, the meaning of a derived word is less unpredictable than the meaning of a compound" (p. 177).

3.2. Idiomatic derivatives outside the phraseological literature

References to idiomatic derivatives scattered in linguistic literature range from an occasional remark to a more systematic type of treatment. Of course, it is not always clear what the authors mean by 'idiomatic' or 'phraseological' and sometimes even the concept of derivative is usually undefined and may be somewhat idiosyncratic. However, it can be expected that the concept is self-explanatory and its meaning is shared by most authors. The mentions are arranged chronologically and they are certainly not meant as an exhaustive inventory of what appears in the literature. Also, the mentions are not restricted only to English.

Igor Mel'čuk (1995), quoted later by Brigitte Horn-Helf (1997: 42) in connection with Russian word-formation says:

> 6. Inclusion of one sign into another with respect both to form and meaning corresponds to ordinary derivation; here three cases have to be distinguished: […] (c) Semi-free derivation: […] in other words, the meaning of the derivative contains that of only one of its formal components plus some quite new and unpredictable 'piece' of meaning, which thus leads to *semi-idiomatic derivatives* [emphasis added] such as *veter* 'wind' – *vetrjak* 'windmill'. In principle, such derivatives arise as a result of condensation, or compression, of syntagmas and semantically they correspond precisely to syntagmas."
>
> (Igor Mel'čuk, 1995: 433-434),

Dietrich Kastovsky (1982) distinguishes between *systematic lexicalization*, such as in the regular addition of very general features such as [+PROFESSIONAL] in derivations by means of *-er* (*lecturer*, *reporter*, *writer*), and nonsystematic, i.e. truly idiomatic semantic lexicalization.

Robert Beard (1987) introduces the term *semantic drift* that affects items stored in the lexicon in both systematic and random ways. The drift resulting in semantic irregularity may start

from primary transparent meanings (such as seen in *construction*, *painting*) or an idiomatic meaning may be subsequently attached to the output of a regular process (as in *transmission* "gearbox"). An affected item 'disengages from the productive L-derivation rule which generates it' (p. 26) and becomes listed in the lexicon.

Robert Claiborne (1990: 223) in *The Roots of English* writes: "*Idiomatic derivatives* include PERMIT and COMMIT, entrust ("put") with ("We now commit their bodies to the deep") — the criminal sense is modern; an officer is entrusted with his commission." What he presumably means is that the morphemic structure of the two verbs is no longer transparent (and so non-compositional) and the verb *commit* has moreover acquired a specific meaning also non-deducible from its form. Ackema and Neeleman (2004) explicitly place both (irregular) derivations and compounds in the lexicon:

> Recall that the lexicon is a list of syntactic and morphological irregularities, containing affixes, simplex words, and idiomatic expressions. The latter are combinations of simplex words and/or affixes that have some unpredictable property that must be listed, for instance a noncompositional semantics. Idioms can either be phrases (such as *kick the bucket)* or complex words (such as *blackbird,* which does not refer to just any black bird, or *transmission* when referring to a car part).
>
> (Ackema & Neeleman, 2004: 54)

In fact, the text is full of references to idiomatic complex words (such as synthetic compounds) which are contrasted with non-idiomatic complex lexical items.

Interestingly, Ray Jackendoff (2009: 652) observes that "irregular plurals (*oxen*, *women*, *axes*, etc.) have to be learned individually and therefore have to be stored in the lexicon. Formally, they are semantically and syntactically composite, but phonologically unitary. They are therefore parallel in structure to idioms, which are phonologically and syntactically composite but semantically unitary. We can therefore think of these cases as 'morphological idioms'."

Storage in the lexicon is related to lexicalisation or the acquisition of non-compositional meaning.

Rochelle Lieber (2009: 63) writes:

> Hand in hand with the notion of transparency comes the related notion of **lexicalization**. When derived words take on meanings that are not transparent – that cannot be made up of the sum of their parts – we say that the meaning of the word has become **lexicalized**. Meanings of complex words that are predictable as the sum of their parts are said to be **compositional**. Lexicalized

> words have meanings that are **non-compositional**. So the words *oddity* and *locality* that we looked at above have developed lexicalized or non-compositional meanings. Sometimes the meanings of derived words have drifted so far from their compositional meanings that it's quite difficult to imagine the compositional meaning for them. Consider, for example, the word *transmission*, which denotes a part of a car.
>
> (Rochelle Lieber, 2009: 63)

The concept of semantic transparency in complex words is covered in great detail by Körtvélyessy, Štekauer and Zimmermann (2015, cf. 3.4.2.).

Haspelmath and Sims (2010: 62-3) when describing the concept of morpheme-based lexicon give the following example of a non-compositional, i.e. idiomatic derivative and draw attention to idiomaticity in derivational morphology:

> First, if the lexicon consists primarily of separate morphemes that are combined together to form words, the meaning of a complex word should be equal to the sum of the meanings of its component morphemes. Stated differently, the word should exhibit **compositional meaning**. But as we have already seen, this kind of direct relationship between form and meaning does not always occur, and derivational morphology presents a particular problem in this regard. A *reader* is not just any person who reads, but also a kind of textbook and the title of an academic job (in the British system). These last two meanings are not predictable from the meanings of *read* and *-er* individually; the meaning is **non-compositional**. This indicates that *reader* (textbook) and *Reader* (British academic title) are probably represented in the lexicon as complex words, rather than according to the component morphemes. The hypothesis that the lexicon consists (almost) exclusively of morphemes thus faces the same practical problem that has led dictionary-makers to give one entry to each lexeme – the meaning of a derived lexeme is often more than the sum of the meanings of the component parts.Since many languages have a large number of derived lexemes with unpredictable meaning, there is correspondingly a significant problem for the hypothesis of a morpheme lexicon.
>
> (Haspelmath & Sims, 2010: 62-3)

In addition, Haspelmath and Sims (p. 95-96) when giving properties which distinguish inflection from derivation claim that unlike inflected word-forms, canonical derived lexemes have non-compositional meaning and "are often semantically idiosyncratic". What is more interesting, the former version of the text (Haspelmath and Sims, 2002) explicitly discusses idiomaticity of derivatives:

> [w]e can distinguish two kinds of idiomaticity. In weak idiomaticity, the semantic contribution of the derivation is present, but the meaning of the derived lexeme is not exhaustively described by the base meaning and the derivational meaning. […] In strong idiomaticity, the regular derivational meaning is not present at all, and the meaning of the derived lexeme cannot even be guessed from the meanings of the components.
>
> (Haspelmath & Sims, 2002: 74-75)

It would be really interesting to find out the reason why the description was changed for the second edition, whether Haspelmath and Sims decided so to avoid the term *idiomatic* in connection with derivatives, or whether this played no role.

Aronoff and Fudeman (2011: 140-141) in *What is Morphology* very simply acknoweldge that "[o]ver time, the meanings of words can become more complex and diverse, making the task of the morphologist looking for semantic patterns of word formation more complicated than it would be if the semantics of word formation were purely compositional (as the semantics of syntactic constructions are often considered to be)." Using the affix *-ism* they demonstrate that it has "some very highly lexicalized meanings", one of which is 'doctrinal system of principles' (*idealism*) and the other an even more specific and lexicalized, describing 'a peculiarity of speech' (*spoonerism*). They conclude that "*-ism* is an example of a suffix with two very highly lexicalized meanings, both of which might be considered to be more characteristic of words than of affixes". Clearly, deciphering the meaning of such words is far from being a straightwordward and compositional process, they are typically listed in the mental lexicon and thus represent "idiomatic" formations.

An occasional reference to an idiomatic derivative appears in Bauer et al. (2013: 30-31): "We should also note that although idiomatization typically occurs with the passage of time, it is nevertheless possible for words to be coined with meanings that are idiomatic from their inception; for example, according to the *OED,* the verb *cannibalize* was attested from the very beginning with the meaning 'to take parts from one machine to use in another'. It has never had the compositional meaning 'to act like a cannibal'."

Needless to say that idiomatic derivatives are reported in other languages as well. Thus Wolfgang Dressler (1994) in *Grammar and pragmatics* has this to say about idiomatic diminutive derivatives:

> Many Italian diminutives that change gender, change other head features as well. Many, but not all, are lexicalized, i.e. idiomatic in meaning. In other words, opacity with regard to the semantics of the base is often connected with opacity with regard to properties of headhood. For example *la finestra* 'the window' has a transparent diminutive *la finestrina,* but an opaque diminutive *il finestr-ino* 'the window of a car/train, on a TV/computer screen', thus no longer a prototypical window (sc. window of a building): this represents a change of a lexical, denotative feature only. The same holds for *la porta* 'the door' —> transparent *la port-ic-ina,* opaque *il port-ello.*
>
> (Dressler, 1994: 105)

Reet Kasik (1997: 42) in his *Typology of Estonian and Finnish Word-formation* makes the following distinction in the Estonian derivational system:

> One part of word-formation is grammatical - words are formed regularly according to the derivational patterns and the meaning of a derivative is determined by a derivational pattern. The derivational meaning of such a derivative constitutes at the same time its lexical meaning […] . The other part of word derivation is lexical - an affix with a certain categorial meaning can form derivatives, where the lexical meaning has become idiomatized and does not coincide with the derivative meaning that is determined by the derivational pattern. Idiomatic derivatives may have varied semantics. The lexical meaning may have concretized in comparison with the derivational meaning. Such idiomatization may involve a whole set of derivatives (e.g. the us-derivative *katus* 'roof of the verb *katma* 'to cover' can be accommodated within the framework of the derivational meaning 'covering device', but it has acquired the additional individual meaning 'part of a building'. On the other hand, the meaning of an idiomatized derivative may have a totally individual character, i.e. it may differ from the derivational meaning of its derivational type […].
>
> (Kasik, 1997: 42)

Lewis Glinert (2004: 167) in *The Grammar of Modern Hebrew*: "Quite distinct from such 'open' verb/adjective + object constructions are 'construct' phrases (described in 6.19), as in: […] These are of three general types. (a) Sometimes they have no 'open' equivalent, eg […] 'colour blind' […]. (b) Often they are *idiomatic derivatives* of 'quasi-object' […] *be-,* eg. […] 'oil- rich', […] 'fear-struck' […] (c) Rarely, they derive from another open equivalent […]".

In a study of verbal morphology of Totonacan, David Beck and Igor Mel´čuk (2011) distinguish several types of phrasemes, among them compound and derivational ones both compooistional and noncompositional. They divide the latter into weak idioms, semi-idioms and strong idioms.

Besides these standard uses of the concept "idiomatic derivative" one may encounter less typical interpretations of the term derivative, which I mention for the sake of completeness. Ermakova et al. (2015) in *Derivation and the Derivational Space in Phraseology as a*

Problem of the Language Contemporary Development provide one example of an idiosyncratic concept of derivation (p. 338) write that "In phraseology the derivation is a processes of formation phraseological units (phraseological derivatives) on the base of phraseological units already existing in a language which are considered to be original and have formal and semantic changes in the derived units' nature. In the derived phraseological units, formed on the base of already existing phraseological units, the structural changes expressed by formal means take place." Some of the examples the authors give, such as 'dojnaja korova' – 'dojnaja korov-k-a', 'zabludšaja ovca' – 'zabludšaja ov-ečk-a' actually do involve a derivational (diminutive) affix, but the 'derivation' the authors have in mind concerns standard multi-word phraseological units.

Finally, often quoted instances of "idiomatic derivatives", which extend the concept of derivative even further, come from the area of sign language and concern the use of one sign to denote a related concept. Thus Klima and Bellugi (1979, 397) write: "What we call here idiomatic derivatives may turn out to be derivational processes with limited productivity. For example, we now find that the process by which CHURCH takes on the meaning 'become narrow-minded' also changes BUSINESS to 'proper.' " Although they are not exactly relevant to the topic at hand, they show an interesting shift in meaning which is not unlike what happens with idiomatic derivatives in the true morphological sense.

3.3. Idiomatic compounds outside the phraseological literature

Like the references to idiomatic derivations, mentions of idiomatic compounds come from authors of all kinds of theoretical background and are presented in theoretical settings of varying sophistication. The main aim of this somewhat desultory collections of quotes is to show that the idea that some compounds are idioms is very common in the linguistic literature and that authors regarded this as a fact taken for granted without, however, following it to its logical consequences.

Richard Beard (1977, cf. Benczes 2006: 22) who considers "semi-idiomatic" compounds such as *blackberry* arguing they fall outside the syntax-oriented framework of generative analysis since their meaning is far more complex than what can be inferred from the meanings of the

constituents and regards them as irregular patterns in the lexicon which should fall outside of transformational rules (or even the synchronic analysis). He holds that the lexical information in the constituents *black* and *berry* is irrelevant and cannot be used to predict the meaning to *blackberry* as such. Such compounds are for him arbitrary, monolexemic linguistic signs that should be assigned to the diachrony of a language and "may explain certain types of jokes, puns and poetry" but are outside of a synchronic study.

In his study on compounds, Jan G. Kooij (1968, cf. Benczes 2006: 79) explicitly points out the existence of so-called idiomatic compounds whose meanings are very often based upon metaphor. He focuses on compounds as endocentric constructions and concludes that although idiomatic compounds initially had the same structures as regular, non-idiomatic ones—they have undergone meaning specialisation to such a degree that they cannot be described by the same set of rules. In other words, they represent a different type in grammar from non-idiomatic compounds. Kooij also observes that the unpredictability of meaning in compounds is not a binary issue, but a matter of degree, which means that decisions between idiomatic and non-idiomatic compounds is ultimately impossible.

Rudolf P. Botha (1968; cf. Benczes 2006: 22-23), who discusses Afrikaans "metaphorical compounds" distinguishes them from "idiomatic compounds", such as *swartbord* ("blackboard"), which refer to a category of linguistic phenomena that constitute a subcategory of the larger category of phenomena referred to by their phrasal correlate (*swart bord*,'a flat slab of wood, that has a colour opposite to white'). This semantic subordination, which Botha calls "narrowing down" of meaning, found in idiomatic compounds becomes evident when the compound and the phrasal equivalent are used in a sentence. While the sentence with the compound *swartbord* is semantically acceptable, the sentence with the phrase is anomalous: **Die swart bord is groen.* *'The black board is green.'

Charles Li (1971, cf. Benczes 2006: 30) explains the acquisition of meaning in what he calls "idiomatic compounds" by "meaning transference rules" and excludes them from his study on the compounding mechanisms of English and Chinese (e.g. *June bug* retains its name regardless of the month in which it appears) in order to limit the types of compounding

mechanisms to a manageable number. He does not offer any method, though, by which idiomatic compounds could be distinguished from "regular" compounds.

An example of a study of nominal compounds within the generativist tradition is Judith Levi (1978), who examines complex nominal with noun–noun nominal compounds (*apple cake, doghouse*) as one subgroup. Levi (1978) bases her theory upon endocentric and partly exocentric compounds. She distinguishes three types of exocentric compounds, those based on synecdoche (e.g. *blockhead* and *cottontail* to describe people and animals respectively, those based on metaphor (e.g. *ladyfinger*, *foxglove*) and coordinated structures (*secretary-treasurer*). She also introduces a "continuum of derivational transparency" (p. 63) for compounds, with completely transparent compounds at one end (*mountain village*), less transparent ones (*briefcase*, *polar bear*), and the third group consisting of exocentric compounds which include the most opaque cases and compounds that are partially or wholly idiomatic (e.g. *flea market* and *honeymoon*).

Pamela Downing (1977: 821) raises an interesting issue concerning Li's idiomatic compounds: if one maintains that any existing compound may serve as a model for the creation of new compounds by analogy, then "one is led to consider the possibility that whole classes of compounds may eventually be derived by analogy to a[n. . .] idiomatic compound". Downing does not follow this idea through and leaves the issue unresolved (the focus of her study was elsewhere). Still, the suggestion that idiomatic compounds might not be so much idiosyncratic after all and might in fact be based on productive mechanisms opens up new theoretical possibilities and has been later taken up by Benczes.

Leonhard Lipka (2002: 95) says: "A complex lexeme may be synchronically analysable but no longer motivated, like *blackboard* or *watchmaker*. If its complete meaning is not derivable from its morphological structure and the pattern exhibited in parallel formations, as in *callgirl, highwayman, streetwalker, pushchair, wheelchair,* we say that such lexical items are idiomatic." Interestingly, in the next paragraph, following Gläser, he claims that "*phraseology* … is concerned with idiomatic *phrasal lexemes* (cf. Gläser 1986: 15). There are also other properties, besides idiomaticity, which distinguish phrasal lexemes from the results of word-

formation and speak for not regarding phraseology as a sub-discipline of lexicology (cf. Gläser 1986:15-25)."

Katamba and Stonham (2006) in their monograph *Morphology* explain the rise of idiomatic compounds in the following way (identifying two causes of idiomaticity, non-adherence to standard rules of word-formation and metaphorical extension):

> However, speakers have the ability to extend the stock of words idiomatically by producing words without meticulously following the standard rules of word-formation. This can be seen in the way in which certain compounds are constructed:
>
> [4.7] a. stool pigeon (police informer)
>
> b. redlegs (poor whites in Tobago)
>
> deadline
>
> No synchronic rules can be devised to account for the meaning of a semantically unpredictable compound like *stool pigeon*. But, in some cases, delving into history might show that some of these compounds originally had a literal meaning that was superseded by later metaphorical extensions. To take one example, during the American Civil War, a *deadline* was the line round the perimeter fence beyond which soldiers were not allowed to go. A soldier who wandered beyond that line risked being shot dead for desertion. (Thankfully, today, going beyond a *deadline* is unlikely to be fatal.) As for *redlegs*, it may be true that poor whites working in the hot sun as labourers on plantations in Tobago did literally have *legs* that were *red*; nevertheless, the compound *redlegs* is semantically opaque. It is very unlikely that anyone could work out the meaning of *redlegs* from the meaning of the words *red* and *leg*. Comparable examples in present-day English are not difficult to find. Consider words like *walkman* and *tallboy*. The former is not a kind of man but miniature personal stereo equipment and the latter is not a boy but a piece of furniture.
>
> (Katamba & Stonham, 2006: 74)

Yet the same authors write later in the book (p. 304): "The nature of idioms will form an important part of our investigations. The reason for this is that idioms raise interesting questions about the interaction between syntax and morphology. Idioms (e.g., *eat humble pie*, i.e., is 'submit to humiliation') are lexical entities and function very much like a single word although they contain several words and are comparable to syntactic phrases or clauses (e.g., [*eat Swiss chocolate*]$_{VP}$)."

In *The Oxford Handbook of Compounding* (Lieber & Štekauer, 2011), Stanislav Kafka devotes one whole chapter to compounds from a phraseological point of view (Compounding and idiomatology: 26-47), in which he "assesses the relationship between compounds and

idioms, arguing that both exhibit a gradience from mildly to wildly idiosyncratic interpretation that begs us to consider them together" (p. 18).

Bauer et al. (2013), quoted above, mention interesting instances of idiomatic compounding in connection with conversion:

> Some examples of conversion from compounds are clearly lexicalized *(to day dream, to pickpocket, to blackmail*), but the examples in (26) show that the process is productive. … The same argument can be made for phrasal elements of certain sorts. For example, we frequently find that phrasal verbs like *blow up, break down, call back, give away, hang out, put down* can appear in nominal contexts. Idiomatic nominal phrases of certain sorts can also appear in obviously verbal environments, as the examples in (27), from COCA, show.
>
> (27) *Fantasy and Science Fiction 199s:* Behind Zane Gerard, Tyque Raymond was **thumb-upping** me.
>
> *Fantasy and Science Fiction 2002:* The cameras were all installed to monitor the reactor, so they faced the center of the room. Most of them **close-upped** on specific pieces of equipment."
>
> (Bauer et al., 2013: 561)

Compositionality is often identified with semantic transparency and non-compositionality or idiomaticity with opaqueness. Borgwaldt and Lüttenberg (2010) understand the former as strength of the relationship holding between the meaning of the constituents of a compound and its meaning as a whole:

> If the meaning of a compound is clearly related to the meaning of its constituents, such as *snowball* is to *snow* and *ball*, the compound is (semantically) transparent. If the meaning of a compound is not clearly related to the meaning of its constituents as in *pineapple* and *pine* or *apple*, the compound is (semantically) opaque. If the meaning of a compound is only clearly related to one of its two constituents, as in *strawberry* (*berry*) or *jailbird* (*bird*), the compound is partially transparent.
>
> (Borgwaldt & Lüttenberg, 2010)

Leah S. Bauke (2017) in her paper 'Content matching in idioms and compounds: a comparative analysis' is quite unambiguous about the existence of idiomaticity at the level of compounds, not only in English:

> It is generally known that certain compound types have an idiomatic reading, i.e. English *redneck, pickpocket, egghead, greenback, walkman, sit-in …* Characteristic of all these is that their meaning cannot be compositionally derived from the meaning of the component parts, nor can the syntactic category of the compound necessarily be derived from the categories of the component parts: e.g. $[sit_V\ in_P]_N$. The same can be found in many other languages (cf. e.g. the examples in (1) in Chinese from Zhang 2007).
>
> (Leah S. Bauke, 2017)

3.3.1. Idiomatic compounds from the cognitive perspective

One of the few authors who analyse compounding from the cognitive perspective is Réka Benczes (2005, 2006 & 2015), who deals with noun + noun compounds and bases her analysis on the theory of conceptual metaphor and metonymy (Lakoff & Johnson, 1980). She emphasizes that "[metaphor] is conceptual in nature: that is, it is not a figure of speech, nor a rhetorical device, but is pervasive both in thought and everyday speech — metaphor is the understanding of one conceptual domain in terms of another" (Benczes, 2006: 89). According to Benczes (2005:195) "metaphor- (and/or metonymy-) based compounds such as *red tape* are not semantically opaque, but can be systematically analysed with the help of cognitive linguistic tools such as conceptual metaphor and metonymy".

Benczes (2006) describes a whole range of semantic types of compounds based on metaphorical or metonymical relations. The basic constituents of compounds are here called profile determinant (i.e. head of the compound) and modifier:

(1) compounds with a metaphor-based modifier: *heartland* "the central part of a country or land". These are according to the author not very common. In addition, there are instances such as *armchair*, which can be understood either metonymically (*arm* stands for *resting of arms*) or metaphorically (*armrest* is like *arm* and *chair* is like *person*) (p. 91-92).

(2) compounds with a metaphor-based profile determinant: *jailbird* "person serving a prison sentence". In this example, *bird* stands metaphorically for *person* based on the association of constraints common to caged birds and jailed people (p. 97).

(3) both elements are metaphor-based: *flame sandwich* "a note that consists of a negative comment surrounded by two positive comments", in which *sandwich* stands metaphorically for a line of comments and *flame* stands for a negative comment in accordance with the conceptual metaphor argument is fire. (p. 103)

Apart from these three types of compounds where one or both elements are metaphor-based, there is also another type in which "the first constituent represents the source domain,

while the second constituent represents the target domain of the metaphorical

relationship" (p. 140):

(4) metaphor-based semantic relation between the constituents of the compound: *bar-code hairstyle* "a style in which a man's last few strands of hair are combed across the top of his head, thus resembling a bar-code pattern". In this example, "the second constituent, *hairstyle*, is the entity that is metaphorically understood as the first constituent of the compound, *bar-code*." (p. 110). Benczes then presents several different subtypes of this type, but the above-mentioned example is sufficient for the present paper.

A similar set of types is presented for the relationship of metonymy (e.g. metonymy-based modifier in *office-park dad* "a married, suburban father who works in a whitecollar job", or metonymy-based profile determinant in *handwriting* "a piece of writing done by hand", in which the profile determinant is based on metonymical relation *action* for *result*).

Four types of compounds which combine both metaphor and metonymy are presented in Benczes (2005 and 2006) and it is claimed that they show some pattern which is considered to go against their interpretation as semantically opaque words. These categories include:

(i) metaphor-based semantic relationship between the constituents of the compound and metonymy-based modifier: *waitress mum* "a woman who is married, has children, works in a low-income job, and has little formal education". In this example, the modifier *waitress* stands for the whole class of women working in low-income jobs. According to Benczes (2005: 188), *waitress* stands for the whole class (and not e.g. *hairdresser* or *secretary*) because it represents all prototypical attributes of the category. As far as the metaphor-based relationship is concerned, Benczes (p. 189) claims that "*waitress mom* is also defined on the background of

the housewife mother stereotype, as a less prototypical member of the mother category, as she does not stay at home with her children but goes out to work instead".

(ii) metaphor-based semantic relationship between the constituents of the compound and metonymy-based profile determinant: *hammerhead* "a stubborn person". In this example, the determinant (*head*) is metaphorically linked to the modifier, i.e. "a head is like a hammer: hard, clumsy, unyielding" (p. 190) and at the same time *head* stands for *person*, based on metonymy.

(iii) metonymy-based modifier and metaphor-based profile determinant: gutter bunny "a person who commutes to work on a bicycle", where *gutter* stands metaphorically for urban neighbourhood roads (which have usually gutters on their sides) and *bunny* is a metaphor *animal* for *person* and is based on the association of swiftness (p. 193).

(iv) metaphor-based modifier and metonymy-based profile determinant: *acidhead* "LSD user", where *acid* is a metaphor for LSD and *head* is in metonymical relationship to *user*.

Benczes is also interested in compounds, which are since Bloomfield (1933) called exocentric compounds. Unlike endocentric compounds, which are described as hyponyms of their head element, exocentric compounds usually include some kind of metaphor or metonymy and they are not hyponymous to their head elements. In accordance with Štekauer (1998: 147), Benczes (2015) claims that exocentric compounds can be dealt with within the same theoretical framework as endocentric compounds, as they are only less prototypical instances of the same relations as (more prototypical) endocentric compounds.

To summarize, in her book on creative compounding in English Réka Benczes (2006) is largely concerned with the possible cognitive processes that underlie noun–noun combinations whose meanings are influenced by metaphor and metonymy. When developing her hypothesis, she has to deal with idiomatic compounds, among other things (p. 12) "by accounting the various ways cognitive linguists have treated the problem of so-called exocentric compounds (or metaphorical, or idiomatic compounds — just to give a couple of examples of the various names by which these constructions have been called) with more or less success." Her treatment of idiomaticity in compounds is different: she subsumes both endocentric and (idiomatic) exocentric compounds under the concept "creative compound" which refers to metaphorical and metonymical compounds alike since as she points out even

endocentric compounds (*armchair, handwriting*) very much like the exocentric *hammerhead* are creative compounds that involve metaphor and metonymy and require the use of creative imaginative, associative processes to be understood. However, these three compounds differ in the degree of creativity they involve and in addition to being "lexicalised to various degrees, a noun–noun combination such as *hammerhead* can be considered to be more creative than *armchair* or *handwriting* in the sense that a greater effort is required from the listener to understand its meaning" (pp. 187-8). In short, what others call idiomatic compounds Benczés regards as a part of the spectrum of creative compounds which (p. 184) "are not unanalysable, nor semantically opaque: in fact, they can be analysed within a cognitive linguistic framework, by the combined application of metaphor, metonymy, blending, profile determinacy and schema theory".

Although Benczes presents a different approach and does not question the status of single-word lexemes within or outside phraseology, it is evident from her work that instances of metaphor and metonymy in complex words are by far no exceptions.

3.4. Indirect description of lexical idiomaticity: meaning predictability

The present chapter briefly describes a linguistic discipline, which is usually understood within the framework of lexical semantics and word-formation: meaning predictability. However, the topic of meaning predictability is also very close to phraseology, since one of the main foci is the relation between the unitary meaning of the whole unit and the meaning of the components in the complex word. In addition, I have decided to include the account of meaning predictability in this study to support my argumentation for the inclusion of lexical idioms in phraseology. The main part of this section will draw on Štekauer (2005a, b), but the topic of meaning predictability in complex lexemes was presented first by Dokulil (1978). Both Dokulil and Štekauer represent the onomasiological approach to word-formation. Onomasiology is a theory of naming in which the linguist examines motivation and possible means and processes leading to the choice of a particular form (expression) for the intended meaning. In other words, in the onomasiological approach, we proceed from meaning to form,

as opposed to the semasiological approach, in which we proceed from the form and seek to establish the various meanings linked to this form. It must be emphasized that the prediction of meaning in fact goes in the semasiological direction, i.e. from the form (perceived by the listener) to the meaning (decoded by the listener). Nevertheless, the onomasiological theory of word-formation has its own apparatus of meaning structure description and this apparatus is used by both authors also in the field of meaning predictability.

3.4.1. Dokulil's concept of onomasiology

Dokulil (1978) focuses on the relationship between the lexical meaning of a word and its word-formation meaning (the meaning of its components). He makes an interesting point when discussing the small interest in this topic, assuming that it is caused by the fact that the topic lies in between two disciplines, word-formation, which has traditionally been part of the study of grammar, and lexical semantics (p. 244). These thoughts show that although Dokulil does not consider the issue to be in the scope of phraseology, he is aware of its unsatisfactory status in both word-formation and lexical semantics.

Dokulil presents three major kinds of relationship according to Igor Němec (1968) and exemplifies them:

1. Full correspondence: the lexical meaning fully corresponds to the word-formation meaning, e.g. possessive adjectives *bratrův* "brother's"*, sestřin* "sister's".
2. Inclusion: a) the lexical meaning is included in the word-formation meaning. This is exemplified by *zelenina* "vegetable", where the adjectival base means "green" and thus the lexical meaning of the lexeme contains the general meaning of "something characterized by its green colour" , but the meaning also contains a specification to "green parts of plants used as a food" (cf. *modřina* "bruise", with the base meaning "blue"*, šedina* "a grey hair") or b) the word-formation meaning is included in the lexical meaning, e.g. *lžičník,* which includes only spoons in the word-formation meaning, but the lexical meaning is "a piece of furniture used for cutlery".
3. Overlap: lexical meaning and word-formation meaning partly correspond, but some elements of the meaning are only part of the lexical and others of the word-

formation meaning. For instance, *truhlář* "joiner" *(truhla* "(wooden) chest" + agentive affix *-ář*) includes in the word-formation meaning not just the activity of making, but also other relations to the original noun – selling, buying, repairing, etc., the lexical meaning, on the other hand, included not just chests, but also other types of wooden furniture). The English equivalent, *joiner*, is a similar case. Apart from its unpredictable lexical meaning "carpenter" it has the regular meaning of a person or thing that joins and the less predictable meaning of a person who belongs to many clubs, associations, societies, etc.

Dokulil investigates which factors influence the relationship between word-formation and lexical meaning. The first factor mentioned is the onomasiological category. The onomasiological category is a term used by Dokulil for "different types of structuring the concept in view of its expression in the given language, i.e., the essential conceptual structures establishing the basis for the act of naming", to quote Štekauer (2005b: 210). Dokulil distinguishes three onomasiological categories: mutational, transpositional, modificational. The modificational type, which is based on adding a modifying feature (e.g. diminutives, augmentatives, change of gender), is basically characterized by correspondence between the lexical and word-formation meaning (*stromek* "a small tree"). Similarly, the transpositional type, where a mark of the base word is transposed into another category (e.g. *hoření* "burning" from *hořet* "to burn"), shows basically correspondence between the lexical and word-formation meaning. In contrast, the mutational type, in which the object of one conceptual category is characterized by its relation to an object of the same or some other category and, in addition to instances of correspondence, also includes instances of other types.

The second factor is the categorial nature which is reflected in the word-class of the naming unit. Nouns show usually the highest degree of specification of the lexical meaning in relation to the word-formation meaning.

The third factor is the word-class and the semantic category of the base word. If, for example, the base word is an adjective expressing a permanent quality, then the verb based on this adjective will show correspondence between lexical and word-formation meaning (e.g. *bledý*

– *blednout*, "pale-to pale"), whereas a noun based on this adjective may have a wider word-formation meaning and a narrower lexical meaning.

The last factor is the word-formation category of the word. In general, derived words in Czech express a lower degree of explicitness than compounds (cf. e.g. *rychlík – rychlovlak,* "express train"). Other factors include clarity/ambiguity of the word-formation structure, distinctiveness/neutrality of the word-formation meaning, the degree of productivity of the word-formation model and the number of instances of its use.

In conclusion, it must be emphasized that the account of meaning predictability is not identical with the account of idiomaticity, in particular, we cannot say that a word with low meaning predictability is necessarily also idiomatic in the prototypical sense. A vague word-formation meaning may be hardly predictable, but it is questionable whether these formations should also be considered idiomatic (e.g. *airer* "a frame for drying clothes on**").**

3.4.2. Štekauer's onomasiological model of word-formation

Pavol Štekauer (2005a, b) uses the onomasiological perspective to define several types of complex words based on the meaning of their parts and studies meaning predictability of actual words within these types. The section will focus on the basic principles of his onomasiological theory and on the meaning predictability of complex words.

Štekauer's onomasiological model of word-formation includes the following levels:

1) Extra-linguistic reality
2) Speech community
3) Conceptual level
4) Semantic level
5) Onomasiological level
6) Onomatological level
7) Phonological level

Extra-linguistic reality represents an object which is to be named. The speech community plays a role as the entity which defines the needs and aims of the communication process. The

conceptual level of the naming process is defined by Štekauer (p. 46) as "a supralinguistic level" which is "independent of any particular language, and represents intellectual processing of the object to be named in a 'coiner's' consciousness by means of generalisation and abstraction processes." The conceptual level includes general conceptual categories, such as SUBSTANCE, ACTION, QUALITY, CIRCUMSTANCE). The semantic level is represented by semes – smallest units of meaning, such as material, animate, human, adult, etc. The onomasiological level is the level of naming in the abstract sense. An important part of Štekauer's onomasiological theory of word-formation is the structure of the naming unit, according to which a naming unit has two basic parts: the onomasiological base (head) and the onomasiological mark which can be complex, i.e. it may include the determining constituent and the determined constituent. At the onomatological level, concrete morphemes are assigned to semes according to the abstract onomasiological structure based on the Morpheme-to-Seme-Assignment Principle (MSAP). At the phonological level, the new unit is modified based on the relevant phonological rules. As was mentioned above, the process of naming and the process of meaning prediction according to Štekauer, are inverted and therefore the levels of the onomasiological model are processed in the reverse order in the process of meaning prediction: starting with the phonological structure and arriving at the referent in the extra-linguistic reality.

However, Štekauer (p. 79) also claims that the extra-linguistic knowledge, reflected in the conceptual level "is involved in meaning identification as early as the onomasiological level and participates in all subsequent steps and related decisions."

1.1.1.1. The onomasiological level

Based on the onomasiological structure, Körtvélyessy, Štekauer and Zimmermann (2015: 86) defines the terms semantic transparency and formal economy. The terms describe two competing tendencies and are of scalar nature: the more a word is transparent, the less it is economic. Based on the abstract onomasiological level, Štekauer distinguishes six onomasiological types. Each type is characterized by its prototypical degree of semantic transparency. The authors (p. 91) claim that that "the central role in terms of semantic

transparency is played by the determined constituent of the onomasiological structure because it is this component that identifies the actual, coiner-determined relation".

In onomasiological type 1, all three constituents of the onomasiological structure are expressed, and therefore, it displays high semantic transparency. An example provided by Štekauer is the synthetic compound *piano-player*, which includes the onomasiological base (AGENT: *-er*), the determined part of the onomasiological mark (ACTION: *play*), and the determining part of the onomasiological mark (OBJECT: *piano*). Onomasiological type 2 combines the onomasiological base with the determined constituent of the mark. It is therefore less transparent than type 1, but still fairly transparent since the central element (the determined constituent of the mark) is still present. Štekauer notes that "[t]he absence of the determining constituent of the mark […] makes complex words of this type more general than those belonging to Onomasiological type 1." (p. 95). An example provided by the author is *teacher*, including the onomasiological base (AGENT: *-er*) and the determined constituent of the mark (ACTION: *teach*), but not including the determining part of the onomasiological mark (i.e. "the person affected by teaching"). Onomasiological type 3 combines the onomasiological base with the determining constituent of the mark. It is therefore less transparent than type 2 because it does not include the central element. An example provided by the author is *bedroom*, which cannot be defined unambiguously based on the meaning of its parts because the relation of *bed* to *room* is not stated explicitly. Onomasiological type 4 includes instances of conversion ACTION-TO-SUBSTANCE. An example provided by the author is *cheat* converted from the verb to noun. It is the most formally economical type, but it is more semantically transparent, because one single morpheme simultaneously represents two onomasiological constituents. Onomasiological type 5 includes the above mentioned joint element of ACTION-TO-SUBSTANCE conversion, but it includes also the determining part of the mark, which makes the type even more transparent. An example of this type provided by Štekauer is the compound noun *miracle-hope* "a person who believes in miracles". Onomasiological type 6 describes structures with two bases are it is represented by copulative compounds such as *actor-manager*. The structure is semantically fully transparent and non-economical.

1.1.1.2. The semantic level

The level of semes is also very important because based on the level of semes, it may be possible to make generalizations about the degree of transparency of words independently of particular speakers. Štekauer (2005a:63) claims that the crucial aspect related to meaning predictability is the seme level of the base because "the onomasiological base (head) determines the grammatical and the lexical features (word class, lexical class) of naming units". An important aspect of the semantic level is the figurativeness of a constituent in the WF structure. Figurativeness of an element lowers the degree of predictability of a complex word. However, it may be still high if the metaphorical meaning is well established in vocabulary.

Štekauer (p. 68) presents his classification of semes into five levels:

1) general conceptual categories (SUBSTANCE, ACTION, QUALITY, CIRCUMSTANCE)

2) classification semes (Animate, Action, Process, State, Quality, Abstract, Tangible, etc.)

3) identification semes (Human, Animal, Plant, Material, Foodstuff, etc.)

4) prototypical semes (Male, Female, Adult, Characteristic material, Characteristic colour, etc.)

5) idiosyncratic semes

The degree of generalisation lowers with each level, with the level 1 semes being most general and level 5 semes being only linked to an individual usage and having unexpected meaning. The most important level in terms of meaning predictability is level 4. In other words, if it is the level four semes which are combined in the meaning of a complex word, the meaning can be considered highly predictable.

Relating the onomasiological theory to the topic of my study, it can be summarized that it is the onomasiological, onomatological and semantic levels which are the most relevant because they describe semes present in complex words, their relation and structure and the concrete realization of the structure.

3.5. Arguments in support of lexical idioms

Both the review of the non-phraseological literature in which authors mention idiomatic derivatives and compounds and the studies of meaning predictability involving indirect

description of lexical idiomaticity make a case for the place of complex lexemes in phraseology. Although the above survey of references to lexical idioms cropping up in linguistic texts not directly concerned with phraseology is very selective and limited (it would be possible to multiply the references indefinitely), it shows clearly enough that there is widespread acceptance of idiomatic derivatives and compounds. Authors have apparently no problem assigning idiomaticity to complex monolexical units, while phraseologists (and sometimes the same authors who have mentioned idiomatic compounds – see Lipka, Katamba & Stonham above) consider it canonical to define phraseology as dealing (only and exclusively) with polylexical units. Obviously, there is a discrepancy, if not a paradox here: how can we say that a compound or a derivative is idiomatic and at the same time exclude it from phraseology? The explanation may be that defining phraseology purely in terms of polylexicality simplifies the delimitation of the field, making it neat and tidy, while the introduction of the lexical level, by contrast, introduces complications. Also the momentum of the entrenched view of phraseology as the study of polylexical units is not easy to shift.

From the above examples of idiomatic derivatives and compounds as described by authors not immediately involved with phraseology it follows that it is especially morphology and word-formation are areas where idiomaticity is frequently noticed. The range of areas bordering on phraseology is, however, wider, as was concluded above in 2.3.

Interestingly, there are not only authors who relate complex words to phraseology, but one author even takes an opposite perspective. Smirnitsky (1998: 210), drawing on his concept of phraseological units as (more or less) word equivalents, attempts to extend this word-equivalence analogy and seeks to find structural and semantic parallels between complex words and multi-word phraseological units. Thus phraseological units with two or more constituents with full semantic value (semantic centres) are for him similar to compounds and their constituents, while phraseological units where only one constituent has full value and the other one is semantically dependent (e.g. phrasal verbs such as *to give up*) are analogous to derivatives consisting of base and affix.

4. Lexical idioms as part of the phraseological theory

As was already mentioned above, Čermák is probably one of the few phraseologists, if not the only one, who explicitly integrates lexical idioms in his classification of phrasemes. Čermák's analyses are mostly carried out on Czech (although 2007b provides examples of lexical idioms also from other languages) and his theoretical work serves as the basis for this study, but as his description is geared towards Czech one of the aims of the dissertation is to inspect which aspects of lexical idiomaticity need to be changed to adapt the theory to the English language.

Like other types of idioms (cf. 2.4.4.), lexical idioms are described by Čermák as units with some kind of semantic, formal or combinatorial anomaly which in their case is restricted to the level of morpheme (or base) combinations. Čermák (2007a: 263) claims that it is more problematic to study lexical idioms in some languages than in others, exemplifying his claim with English in which there is a number of polysemous suffixes with a wide range of meanings, such as *-er* (which may signify the doer, instrument, result and have many other meanings). In addition, there are words with identical strings at the end which, however, cannot be assigned to one and the same group. These strings may represent a homonymous suffix (*redder*), they may be part of a monomorphemic word (*power*) or they may be of a questionable status (*anger*).

4.1. Morphological classification of lexical phrasemes

In his comprehensive work on Czech and general phraseology, Čermák (2007a: 264) classifies Czech lexical phrasemes (calling them phrasemes in this respect, because of the focus on form) according to their morphological structure; in particular, the word-formation process involved and the resultant word-class. Accordingly, he distinguishes two principal groups, derivational and compositional phrasemes, subdividing each of the two into four main classes: nominal, adjectival, verbal and adverbial. Lexical phrasemes (that is word-level phrasemes) can be found in other word-classes too, i.e. functioning as grammatical word

classes (prepositions, conjunctions) and discourse markers or particles, but the above four classes are most prominent.

Figure 7 illustrates this classification with examples from Czech:

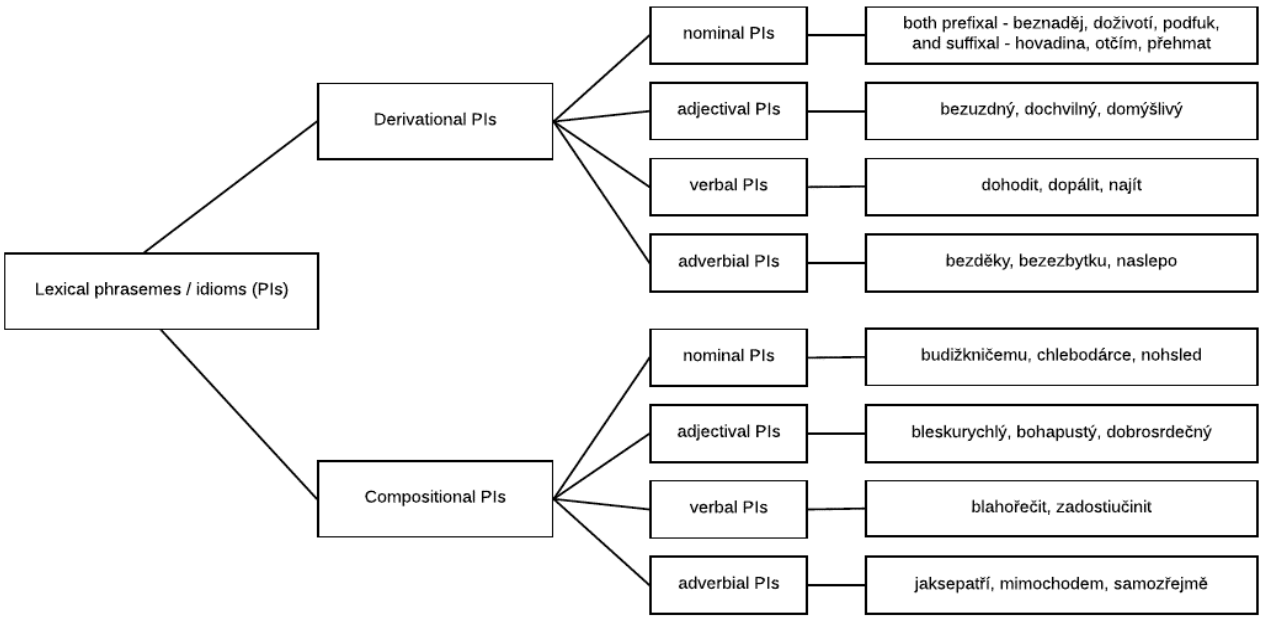

Figure 7: Morphological classification of lexical phrasemes according to Čermák (2007a)

Čermák identifies several subtypes of idiomatic derivatives. Firstly, there are idioms containing monocollocational elements (elements which can be found nowhere else in the language, e.g. *-ím* in *otčím,* "stepfather"). There are a few examples of a circumfix (*do-život-í* "life sentence") and the largest group comprises instances of common prefixation or suffixation. Probably the most numerous group is, according to Čermák, that of verbal derivational lexical phrasemes. Some verbs become idiomatic through a change in valency or reflexivity (e.g. *vzpomínat / vzpomínat na + ACC, nedat / nedat se*). Čermák (p. 267) includes them in the category of lexical idioms, but stresses that they are borderline cases between lexical and collocational idioms.

Another approach to the morphological study of lexical idioms presented by Čermák (2007a: 268) is the analysis of all possible prefixes and suffixes used with a base containing a common verb (*e.g. být,* "to be", *jít*, "to go"). Using this approach it is possible to find the ratio of regular and irregular (idiomatic) combinations.

Čermák (2007b: 24) also claims that languages which prefer derivation to compounding, such as Czech, provide good examples of lexical idioms. He mentions an example of idiomatic prefixed verbs, which can be subclassified according to the type of anomaly into these three classes: a) pure additive prefixation (*vyjít*), b) prefixation accompanied with a change in valency (valency removed, *přebrat;* valency substituted by a different case, *zahýbat někomu;* valency added, *vyzrát na*), c) reflexive pronoun added (*zašít se*).

4.2. Semantic types of lexical idioms

Čermák (2007a: 273) also studies semantic types of lexical idioms and finds a significant group of evaluative words (e.g. *kravina, podfuk, dobromyslný*). Another, related, point is the high incidence of the negative prefix *ne-* among lexical idioms (these lexemes are also often evaluative). A significant group from the semantic point of view is that of instances of folk terminology (cf. 2.9.). These are morphologically complex words based on some idiomatic relation between elements which have lost their vagueness and attained an exact definition over time. This moves them closer to regular language. The majority of lexical idioms refer to abstract concepts, which is, in Čermák's view, in accordance with other types of phraseological units.

4.3. Problems with the identification of lexical idioms

There are two main problems with the identification of lexical idioms which are addressed by F. Čermák (2007a: 264-265). Firstly, it is rather difficult to draw a clear line between the assumed usual meaning of a given morpheme and its idiomatic use. Čermák considers the most frequent meaning found in the dictionary to be the prototypical one. However, this approach will leave a lot of analysed words somewhere on the borderline between regular and idiomatic. In other words, we can still feel some regularity even in combinations whose parts have a less frequent meaning. Čermák also explains that he does not include such words as *jackpot* among lexical idioms referring to such verbs as semantically depleted, claiming:

> [t]he crux of the matter is in the difficulty to discern between a very large number of meanings (dictionary senses) and a loss of meaning (depletion). In this light, *jackpot* (with 14 meanings for *jack* and 6 for *pot*, according to "The New English Dictionary of English") represents something of a borderline case, since the number of meanings for *jack* is just too large and, primarily, wide-ranging. It seems that this type of coinage should refer to a basic, prototypical meaning of the constituent if viewed independently, which however is rather difficult to find for *jackpot*
>
> (Čermák, 2007b: 23-34).

Another problem, which partly overlaps with the previous one, is the degree to which we will take into account the diachronic perspective. Čermák claims that synchronic analysis has only a limited value as it is problematic to distinguish between the synchronic and diachronic relations without complete data on the whole vocabulary. Čermák (2007a: 265) therefore concludes that we need to base our analysis on intuition and estimation relying on synchronic data which does not make the procedure any easier to apply.

4.4. A quantitative study of lexical idioms in Czech

While Čermák focuses on the theoretical description of lexical idioms, Klötzerová (1997, 1998) conducted a study of Czech lexical idioms based on data retrieved from a Czech dictionary. Klötzerová bases her analysis mostly on the concept of phraseology developed by Čermák, for whom the basic property of (lexical) idioms is their (multiple) anomaly. The presence of anomaly is considered to be the main criterion for idiomaticity. The features taken into account are the range of paradigm, productivity, and the substitution and transformation tests. These tests are applied in accordance with Čermák's definitions. Idioms with anomalous valency are assigned by Klötzerová to the category of lexical idioms because valency is a grammatical property of lexical units and a regular lexical unit can combine quite freely with other elements.

Klötzerová stresses that some aspects of multi-word idioms cannot be applied to lexical idioms. Unlike some multi-word phrasemes lexical idioms exhibit lexicalization (i.e., all lexical idioms are lexicalised) and syntactic frozenness (fixed order of elements characterizes all words including lexical idioms). Lexical idioms are then identified on the basis of two features: morphological complexity (they are polymorphemic words) and multiple anomaly.

Anomaly can be formal, collocational and semantic. Formal anomaly means that the complex lexeme is formed by a non-productive process. Collocational anomaly occurs in words consisting of constituents (morphemes, bases) with limited lexical combinability. Semantically anomalous i.e. non-compositional, are those words in which the total meaning is not the sum of the meanings of their constituents.

Idiomaticity is treated as a graded phenomenon. Klötzerová works with the terms centre and periphery and distinguishes between central and peripheral lexical idioms. Peripheral lexical idioms display fewer features of idiomaticity than central ones. Peripheral lexical idioms include terms, phrasemes with a non-specific component (semantically wide), phrasemes with a monocollocable component, phrasemes semantically transparent, phrasemes with a semantically vague base, phrasemes with a structure synchronically almost opaque.

Klötzerová identifies approximately 5% of lexical units in the Czech dictionary *Slovník spisovné češtiny* (1994) as lexical idioms. Lexical idioms are then categorized according to their nature as compositional idioms, prefixal idioms, valency idioms and reflexive idioms. The largest group is that of prefixal idioms, which is in accordance with the general word-formation pattern of the Czech language. Klötzerová then studies expressivity of lexical idioms and finds that about 24% of them are expressive. Rather frequent are lexical idioms with a concrete denotation (69% of nouns without terms). The group of compositional idioms includes mainly nouns and adjectives of the structural types adjective-noun or noun-verb. Quite common is the process of juxtaposition (*jaksepatří*). She finds that 42% of all noun compounds are personal names which are formally special and expressive. Prefixal idioms are largely represented by verbs. These verbs are mostly idiomatic in their primary sense, unlike valency idioms, which are idiomatic usually in their secondary sense. Valency and reflexive idioms are more often expressive than prefixal and compositional idioms. Specific groups of lexical idioms are phrasal compounds, adverbs of the type *dočista* (completely), lexical phrasemes with very limited collocability, impersonal verbs, polysemous derivatives from verbs of motion, polysemous derivatives bound contextually, derivatives with a semantically vague base, adjectives of the type *zašlý* (faded), and phrasemes with specific etymology.

4.5. Čermák on lexical idioms in English

Apart from Czech examples, Čermák (2007b: 22) also lists some examples from other languages, including English. Among compounds, he mentions e.g. *breakfast, greenhorn* and *man-of-war* with semantic anomaly (non-compositionality), but also *eavesdrop* and *cranberry* with collocational anomaly, containing elements *eaves-* and *cran-* which are not to be found anywhere else in the language. Čermák (p. 23) stresses that "it is a rich field requiring a systematic study and it should not be viewed as limited to metaphors only." A distinct subclass of English lexical idioms is the subclass of particle compounds, e.g. *getaway, comeback, take-off*.

To summarize the final chapter, Čermák is the first one to overtly include lexical idioms in the theory of phrasemes and idioms (with Klötzerová testing and confirming the theory in practice), exemplifying the theory mostly by examples from Czech. The main aspects of lexical idioms studied by Čermák and Klötzerová are the type of anomaly and the semantic class of the lexeme. Čermák's description of various aspects of lexical idioms serves as the main part of the theoretical background of the following study of English lexical idioms, in which I attempt to inspect the plausibility of including lexical idioms into phraseological research and adapt the theory designed for Czech to English.

5. Methodology

Before I proceed to the analysis of English lexical idioms, I will first give a detailed description of the objectives, including research questions, which determine the methodology and the scope of the study. The chapter then presents how the analysing process will be structured to reflect the aims and research questions. The next section focuses on the provisional definition of lexical idioms and also discusses some general issues which may influence the method and the result of sample extraction. The last two sections describe the two compiled samples and reasons why they were built in that particular manner. It must be stressed here that the study takes a synchronic perspective. This is especially reflected in the definition, classification and conclusions which the study arrives at. However, there are instances where the diachronic perspective is also discussed, especially when it is necessary to describe some of the systematic issues of particular classes of lexemes.

5.1. Objectives and questions

As we have seen in the previous chapters, single-word complex lexemes are not traditionally described as falling within the scope of phraseology, although many linguists from various theoretical backgrounds discuss their idiomaticity. Therefore, the most general research question posed in this study is:

1) Is it reasonable to study anomalous combinatorial relations below the level of the word within and by means of phraseology?

Of course, it may be problematic to arrive at a definite answer to such a general and subjective question. Therefore, what the study attempts to do in this respect is to find and describe both possible similarities with traditional polylexical phraseological units and also the dissimilarities. It seems plausible to expect that similarities will support the theory of lexical idioms, whereas dissimilarities will work against the theory and weaken the status of single-word lexemes within phraseology. In order to proceed to the more specific questions, we need to presuppose that the answer to the first question is, at least to some degree, yes. In

fact, the argumentation carried out above, in Chapter 3 and 4, suggests that it is indeed the case.

The second research question concerns the very definition of lexical idioms. As we have seen in Čermák (2007a), he posits three possible sources of anomaly in complex units: semantic, formal and collocational. Other authors mostly mention only the former two anomalies in connection with polylexical phraseological units. They are mostly referred to as semantic non-compositionality and syntactic frozenness. Research question 2 can be formulated as follows:

2) Are all three types of anomaly, i.e. semantic, formal and collocational, of equal importance to the definition of lexical idioms, or is any of them more important for the identification of lexical idioms than the others? Conversely, is any one type of anomaly less important for defining lexical idioms, and should be even disregarded?

Since the only systematic description of lexical idioms within phraseology is carried out on the Czech language, I expect to find differences between the results of the Czech and the English inquiry. These differences may be due to both typological and historical reasons: An important typological difference is that Czech is a language with rich and productive derivation, whereas English is richer than Czech in composition. The historical differences are above all represented by the fact that the proportion of loanwords in English is higher than in Czech and they are mostly well assimilated into English. Moreover, components of these loanwords to a certain degree follow systematic patterns even though the processes are unproductive. This happens in Czech as well, but to a lesser extent. Therefore, an important objective related to the question above is to find a way in which to adapt Čermák's definition of lexical idioms to English.

In addition, there are several minor questions, which are to some degree related to the previous two:

3) Are there any types of lexical idioms which were not described in the literature so far?

4) Are there any formal categories of lexemes which are more typical among lexical idioms?

and conversely:

5) Are there any formal categories of lexemes which are less typical among lexical idioms?

6) Are there any semantic types which are more typical among lexical idioms?

7) Is expressivity a feature which is typical of English lexical idioms?

The last question returns to the traditional description of idiomaticity within the phraseological approach. Assuming that the idiomaticity of single-word lexemes is as scalar a quality as in polylexical units:

8) Are there any specific classes of lexical idioms with respect to the degree of idiomaticity they display?

The answers to the questions are expected to result in a description and overview of the phenomenon which can be further investigated from various perspectives in subsequent studies.

5.2. A two-stage analysis

The present study is carried out in two stages. There are two samples of lexemes extracted from two different sources. The aim of this procedure is to first make a general overview of possible lexical phrasemes in English, testing a provisional set of criteria which is based on the literature, especially on the description of three different kinds of anomaly by Čermák (cf. Chapter 4 above). The sample is comprised of 1 000 random lexemes (see 5.4. for details). Next, on the basis of this general overview, some adaptations are proposed to the theory which should reflect the specific nature of the English language. This step is presented in Chapter 6.

The new criteria are then employed to obtain a new sample of 500 lexical idioms from the OED. In this second sample, I attempt to minimize the problems related to the specific nature of English vocabulary by excluding some problematic and less prototypical categories and instances of lexical idioms. The OED sample of lexical idioms is then analysed in terms of both formal and semantic types of lexemes. In addition, selected pragmatic functions, as one of the features mentioned by Čermák in association with collocational idioms (cf. 2.8.), will be taken account of.

5.3. The provisional definition

The starting point is provided by two general definitions of phrasemes/idioms offered by Čermák (1982: 115-116, English version in 2007: 83). The first definition reads "The idiom and phraseme is a unique combination of minimally two elements, one (or more) of which does not function in the same way in another combination (combinations), or it occurs in just one expression (or a severely limited number of such expressions)." The second definition says that "the phraseme and idiom is such a non-model and fixed syntagma of elements of which (at least) one is with respect to the other a member of an extremely limited (both formally and, mostly, even semantically) and closed paradigm."

Klötzerová (1997: 9) then formulates a definition of lexical phrasemes/idioms based on the above definitions as follows: The lexical phraseme/idiom is a (fixed) combination of at least two morphemes which is anomalous from the formal, collocational and semantic points of view and the result of which is a single-word lexeme.

The definition provided by Klötzerová is probably too strict in emphasizing that all three kinds of anomaly are involved. She does not analyse the types of anomalies for each particular lexeme and does not discuss this topic in the empirical part, and therefore we can only speculate about the specific anomalous features of the studied lexemes. On the other hand, what the definition stresses is that anomalies of different kinds often combine inside one complex lexeme. Therefore, I will base the definition used in this study on the above three definitions, while not insisting on the occurrence of all three types of anomaly at the same time. The definition used in this study is thus as follows:

The lexical idiom is a single-word lexeme formed as a combination of components which is anomalous semantically and/or collocationally and/or grammatically.

The first part of the definition is probably not very problematic: The lexical idiom is a single-word lexeme, as opposed to combinations of words (collocational phrasemes and idioms in Čermák's terms). However, problems may arise even at this point of definition as there are instances of combinations in the fuzzy area between compounds and word combinations (especially open and possibly also hyphenated compounds, e.g. *apartment building, mother-*

in-law, but also complex prepositions, e.g. *due to*). These are dealt with in this study with the help of the standard criteria of wordhood:

1) orthographic criterion: the lexeme is written as one word (or with a hyphen between its parts)

2) phonological criterion: the lexeme has one primary stress, the compound lexeme is forestressed

3) semantic criterion: the lexemes denotes a unitary concept

4) syntactic criterion: the lexeme functions syntactically as one unit: it is normally indivisible and no part can be substituted by another.

It is obvious that none of the criteria is absolute and that the quality of wordhood is scalar and the inclusion into or exclusion from the category of words may be subjective in some borderline cases. Although the borderline between words and collocations is an important area for the present study, I don't think this peripheral area will pose a serious problem as we are attempting at applying the same methods which are normally used for analysis of word-combinations and therefore it is possible to view the categories of lexical and collocational idioms as overlapping categories with no clear-cut borderline between them. In addition, this problem is almost eliminated in the BNC sample because only lexemes written as one unit are extracted (and thus one does not have to deal with the largest fuzzy area, i.e. with open compounds). In the OED sample, where orthographic criterion points to collocation (i.e. the lexeme is written as two separate units), I first use the phonological criterion and then the syntactic and semantic ones.

The next part defines a lexical idiom as a combination of components. By components I mean both single morphemes and polymophematic components (complex bases) participating in the process of further word-formation. In addition to these, I include also splinters, i.e. irregular components in the form of a fraction of a word produced by clipping regardless of the morpheme boundaries, which appear especially in blends (cf. Bauer, Lieber, & Plag, 2013: 525). The reason for including splinters, and thus blends, into the description of lexical idioms is that they carry some meaning which is present in the new formations.

Semantic anomaly as the defining feature of lexical idioms is one of the most problematic features, yet really important. A lexeme is semantically anomalous if its meaning is not the sum of the meanings of its constituents, and vice versa. However, prototypical cases of unequivocal compositionality are extremely rare at the level of words and they may include examples from the onomasiological category of modification (Dokulil, 1978), e.g. *droplet, greenish* and the category of transformation, e.g. *redness, exemplify*. In most cases there is either overlap of the two meanings (see Němec, 1968), e.g. *blackboard* (which does not have to be black and does not have to be a board) or the inclusion of one meaning in another, e.g. *scarecrow* (which does not scare just crows but also other birds). By contrast, instances of complete disjunction occur especially in cases of metaphoric or metonymical shift in meaning, e.g. *egghead* (who is not a head and has nothing to do with eggs).

Therefore, we need to define a borderline between combinations which may be described as (fairly) transparent or compositional and those which are (more or less) opaque and thus semantically anomalous. There are three possible locations of the anomaly in the syntagma: either of the two parts or the way they are combined (cf. Benczes, 2006, or section 3.3.1. above, but also the onomasiological structure by Štekauer, 2005, or section 3.4.2. above). A constituent is considered to be anomalous if it does not occur with the same meaning as the corresponding independent word or in other combinations, or if it occurs in the given meaning only marginally, especially if it has undergone a shift in meaning involving metaphor, metonymy, meaning extension or semantic specification, and other processes. This can be practically tested by the recurrent contrast test described by Cruse (2000: 70), which is also mentioned as the commutation test by Čermák (2007: 83). In addition to the test, we can also study the meaning of the constituents in the dictionary taking into consideration the frequency of the given forms. This can be exemplified by the lexeme *understand*, which can be considered a prototypical instance of a lexical idiom. The first constituent, *under,* when used as an independent word, has the main meaning "in, to or through a position that is below something" and a number of other frequent meanings. The OALD names also "less than; younger than", "used to say who or what controls, governs or manages somebody/something", etc. The prefix *under* is defined as "below; beneath", "lower in age or rank" and "not enough". None of these senses can be linked to *under* in *understand.* The use of the contrast

test shows that there is no such thing as **overstand* (as in *underestimate* and *overestimate*). A similar procedure could be applied to *stand* with a similar result, except for the fact that the contrast test may be harder to carry out in the case of lexical bases because of the lack of possible analogies. On the other hand, if the lexeme *underestimate* were tested, the meaning of both constituents would prove to be analogous to their meaning as independent words and it would pass the contrast test (*underestimate – overestimate; underestimate – underrepresent*).

A type of semantic anomaly closely linked to metaphor and metonymy is exocentricity (cf. also 3.3.1). Exocentric compounds are defined as compounds which cannot be described as hyponyms of their syntactic head and they are traditionally described as a distinct class of compounds. Nevertheless, Bauer (2016: 474) points out that "[m]ost, perhaps all, exocentric compounds can be viewed as figurative uses of endocentric constructions. In particular, even where the head of the compound is not of the same word-class as the compound, it seems that most cases of exocentricity can be viewed as instances of metaphor or metonymy (with synecdochic compounds often singled out for particular mention in the literature as bahuvrihis)." If exocentricity is understood as a type of figurative use, then it must be understood as an important type of semantic anomaly and exocentric formations must be seen as lexical idioms.

If the anomaly does not involve the meaning of one (or both) of the constituents, but consists in the manner they are combined, we must consider it a combination of semantic and collocational anomaly. This kind of anomaly may be recognized especially if the two meanings of the two constituents are incompatible. An example of this phenomenon is the exocentric compound *butterfingers*: unlike premodifying adjectives typically belonging to the class of relational or qualitative adjectives, the attributive *butter* cannot be simply viewed as the material, purpose, type or quality of *fingers.*

Collocational anomaly may be also linked to formal features rather than semantic features. This happens when a constituent of the lexeme combines with a limited number of other lexemes or even with only the one instance. A common example of this type is *cranberry*. However, this class is rather problematic with respect to the present study as the inclusion of

such words among lexical idioms may go against my aim to focus on language from the synchronic point of view as much as possible and many examples of this kind are on the borderline between complex lexemes and simplex, monomorphemic ones.

The most typical examples of grammatical anomalies of multi-word phraseological units cannot be simply transferred to the level of single-word lexemes. Grammatical fixedness, or frozenness, is typical of all lexical units that have undergone the process of lexicalisation and are thus stable in form. Grammatical anomaly in single-word lexemes may be represented by the use of non-productive affixes (cf. Čermák, 2007a: 74), e.g. *a-* in *afloat* or *-ible* in *collectible*. Another example is the instance of partial conversion of adjectives to nouns, e.g. *the poor*, *the mysterious*. However, partial conversion is a less prototypical example as it is questionable whether the combination of a definite article with a monomorphemic word (supposing we understand conversion as an instance of relisting the item into another category, not an instance of zero-derivation) can still be considered a single-word lexeme.

It needs to be emphasized that anomalies of various types may combine within one lexeme as shown by the example of semantically incompatible components in *butterfingers*. However, *butterfingers* can be also seen as an example of a grammatical anomaly, as the word has the plural -s (as part of the second component), added prior to the process of composition, but the word functions as a singular and does not normally form the plural.

5.3.1. Frequency as an indicator of anomaly

It is evident that the crucial part of the sample collection is the identification of anomalies. The basic indicator and criterion used to assess whether a lexeme displays any anomaly in its semantics, collocability or grammar is the frequency of the studied phenomenon (the possible anomaly) within other lexemes in the vocabulary. However, the use of the frequency criterion must be applied somewhat differently for each type of anomaly.

It is relatively easy to recognize a grammatical anomaly: the rules of a grammatical system normally apply to large classes of items while the frequency of an anomaly is considerably lower than the frequency of regular uses of the grammatical rule. This pattern may be more difficult to follow when etymological changes interfere with the structure resulting in various alternations (e.g. base alternation in *pronounce – pronunciation*).

Collocational anomaly poses a problem of a different kind: formal collocability of two components may be less evident and less easy to quantify and we are bound to come to the conclusion that there is a large fuzzy area where it is difficult to decide whether a phenomenon should be viewed merely as less frequent or as anomaly. An example of this is the Romance compositional type *pickpocket*. There is a class of lexemes formed similarly, with the first component being the verb and the second component its object, but the class is significantly less numerous than other nominal compositional types (e.g. the N-N type or the ADJ-N type). It seems therefore plausible to consider this compositional type to be rather anomalous and include it in the study, although the type is less idiomatic than the type V-N with an agentive noun which can be exemplified by the lexeme *crybaby.*

The frequency criterion reveals a large fuzzy area also in the case of semantic anomaly. To assess whether the given sense of a component is anomalous or not, one needs to decide whether the sense is normally associated with the morpheme/component or not, which can be done with the help of a dictionary[2]. The approach used in this paper is that I attempt to assess frequency relatively, in comparison with other competing senses. Unfortunately, frequency lists available for English are normally based on lexemes, not senses, and therefore to obtain objective data about the frequency of senses of a lexeme (or morpheme) is difficult. For this reason, semantic anomaly must be assessed partly by intuition in combination with dictionary data. Any transfer of meaning (based on metaphor or metonymy) or meaning generalization or specification also leads to the inclusion of the lexeme in the sample as these changes are considered an anomaly.

5.4. Collecting the BNC sample

The first analysis is carried out on a sample of 1 000 lexemes. The sample has been extracted from the British National Corpus (BNC), Version 3 (BNC XML Edition). The BNC comprises over 100 million words from approximately 4 000 texts. Broadly, 90% of the corpus are written sources, 10% are spoken texts. The written sources cover non-fictional genres (from 1975 to the early 1990s) and fiction (from 1960 to the early 1990s).

[2] Collins COBUILD Dictionary has proved to be well suited for these checks because of its systematic and concise presentation of lexeme senses.

The aim was to retrieve a sample of random lemmas of all relevant word-classes which would represent the phenomena of compositionality and regularity (and, of course, non-composionality and irregularity) in all its shades. Therefore, I have retrieved a frequency list based on "headword or lemma frequencies", restricting the choice to open word-classes (N, ADJ, V and ADV) with no other restrictions but the lowest frequency of 10 occurrences. The list was then randomized by means of the spreadsheet function and the first 1 000 items were used for the analysis, excluding proper names, initialisms and some other irrelevant items (e.g. numerals tagged as nouns or adjectives, word-like components which only occur as part of collocational phrasemes, etc.). The reason for restricting the sample to open word-classes is that they fall within the scope of productive word-formation rules and we can thus expect to find all degrees of productivity and compositionality among them. Closed word-classes (i.e. functional words), proper names, numerals and initialisms were excluded because they are marginal with respect to the topic of lexical idioms.
The sample collected according to these principles thus contains both simple and complex words and both words formed within English and loanwords.

5.5. Collecting the OED sample

In the second sample, I focus on the more prototypical types of lexemes associated with idiomaticity. Therefore, the basic features of the sample were set before the analysis of this and the first sample, but some specific features will be described later on the basis of the results of the first analysis and the amended definition of lexical idioms.

The source of the second sample is the Oxford English Dictionary (OED), which is the leading historical dictionary of English, containing about 600 000 words.

5.5.1. Differences from the first sample

The main intention for the collection of the second sample was to avoid:

a) structurally problematic lexemes of early origin. By this I mean either etymologically complex lexemes in which the structure is opaque due to linguistic change or unproductive,

opaque affixes (e.g. *among*, originally a phrase *on gemang*, or *around*, *a-*, *round)*, or early loanwords which were formed by word-formation processes in the donor language and are opaque or almost opaque in English (e.g. *involve, require*);

b) lexemes formed outside English and borrowed as one unit, although sometimes more or less analysable. This restriction was included because loanwords present a theoretical problem: the traditional definitions of phraseological units all refer to a combination of components, and there are two possible interpretations of this. The narrow-scope interpretation would only include units formed by combining the two or more components within the given language (i.e. within one particular system of grammatical and semantic word-formation rules), whereas the broad-scope definition would include all units consisting of analysable parts regardless of diachronic aspects. Classes of lexemes which stand at least on the periphery of the studied phenomenon are discussed in the BNC sample, and thus there is no need to focus on them in the OED sample.

The core for the OED sample was accordingly extracted as follows: The "advanced search" in the OED web page offers the possibility to restrict the "date of entry" of a lexeme into the OED. This makes it possible to eliminate words affected to a large extent by linguistic changes over time. The extraction was confined to the period from 1800 up to the present. This period is long enough to give quite a heterogeneous sample of words, but at the same time includes mostly words formed by processes which are still productive at the present time. These data were then sorted according to frequency and then only complex lexemes were manually selected from the dataset. This list of lexemes then served as the basis for the extraction of 500 lexical idioms, the extraction of which is described in detail in section 6.2.1.

5.5.2. Frequency in the OED

As *Key to frequency*[3] describes, "The underlying frequency data is derived primarily from version 2 of the Google Books Ngrams data. This has been cross-checked against data from other corpora, and re-analysed in order to handle homographs and other ambiguities." The data include only non-obsolete words and the frequencies are calculated from the sources

[3] https://public.oed.com/how-to-use-the-oed/key-to-frequency/

dated from 1970 onwards. Lexemes are sorted according to their frequency into 8 bands, from band 8 (very high-frequency words) to 1 (very low-frequency). The frequencies are summarized in the following table:

band	frequency per million words	% of entries in OED
8	> 1 000	0.02%
7	100 – 999	0.18%
6	10 – 99	1%
5	1 – 9.9	4%
4	0.1 – 0.99	11%
3	0.01 – 0.099	20%
2	< 0.0099	45%
1	–	18%

Table 4: Frequency bands in the OED

Although the system of frequency marking is relatively crude, with large bands of units of equal value, it serves well for the purposes of the present study: it filters the lexemes in a way that eliminates rare, strange or exotic words, technical terms from specialized discourses, etc.

5.5.3. Disadvantages of the method

There are certainly also disadvantages caused by the restrictions imposed by the method of sample extraction. The most important disadvantage is that we cannot obtain representative quantitative data on the occurrence of lexical idioms in the whole vocabulary as the sample is significantly restricted in time. This is especially because the lexemes formed within a restricted period of the last 200 years are to some extent homogeneous: the sample includes a strikingly large proportion of terms (although only those in general use due to frequency restriction), names of new concepts from both natural sciences and humanities, i.e. many of the newly formed words are terms and abstract concepts in which the relation of the lexical and word-formation meaning may be difficult to describe. This can be partially compensated

for by excluding terms (which are outside the scope of phraseology anyway, cf. 2.9.), although not completely (see 6.2.6. for details).

It is also questionable whether data sorted by frequency bands do not exclude some distinct classes of lexical idioms. This uncertainty can be partially dealt with by complementing the data with the BNC sample.

6. Analysis

6.1. The BNC sample analysis

6.1.1. Introduction

As was already described above, the sample contains 1 000 randomly selected lexemes of word-classes that can be potential idioms. The aim of the analysis is to map the situation as a whole focusing on both the centre and periphery of the field. The following table shows the distribution of the four word-classes studied and the distribution of simple and complex lexemes:

WORD-CLASS	SIMPLE	COMPLEX	TOTAL
N	206	322	528
ADJ	13	264	277
ADV	2	47	49
V	98	48	146
TOTAL	319	681	1 000

Table 5: Distribution of word-classes and simple/complex lexemes in the BNC sample

Although the differentiation of simple and complex lexemes is only the first step towards the analysis, it is evident already at this stage that we need to define the borderline more precisely. Besides obvious instances of simple words, such as *white, rare, catch* and *money*, and obvious instances of complex words, such as *unrecognizable, nippy, government-owned* and *milkman*, there are also words such as *conflict, prepare, impact, odour* which are placed somewhere in between because they are partly analysable, especially for an English speaker with some knowledge of classical languages. And then, there are lexemes such as *superficial*, *amenity*, but also native *Tuesday* and *beware*, which are more likely to be classified as complex, although it is also problematic to describe their structure from the synchronic point of view. The following section describes the classification based on both form and meaning of the lexeme and shows how some of the specific groups of lexemes were dealt with.

6.1.2. Classification of the sample items

6.1.2.1. Simple lexemes

Simple lexemes are lexemes consisting of a single morpheme. They are outside the limits of phraseology, since we understand phraseological units as combinations of components. However, as was already noted above, the distinction between simple and complex can be seen as a scale. The simple lexemes in the sample can be ordered in groups on a scale from prototypical to borderline cases, or in other words, from the centre of the category to the periphery:

The first group includes unproblematic simple lexemes. The best, and least problematic, examples are monosyllabic. They are mostly native words and well-assimilated early borrowings (ex 1):

(1) *catch, peace, right, queue, mass, white, spit, past, twist*

There are also polysyllabic words which are unequivocally simple. They include partly native words, but also many loan words, both assimilated and unassimilated (ex 2):

(2) *money, fiesta, blossom, abandon, heckle*

Secondly, there are simple borrowed lexemes with some traces of the original complex structure. These lexemes were complex in the source language and speakers, especially those familiar with the source language, can still see some traces of the original structure, although there is no obvious link between the "components" and their meaning. From the synchronic point of view, they are simple in English (ex 3):

(3) *vagabond, instant, anthem, delicatessen, insect, perplex*

The third group contains simple lexemes whose one part accidentally resembles an existing affix (ex 4):

(4) *privy, beaver, gutter, mayor*

The last group consists of complex borrowed lexemes whose components are both formally and semantically recognizable in English. This group is the most peripheral among the simple

lexemes. These lexemes of Romance or Greek origin contain a component which is systematically used in English with the corresponding meaning. However, due to linguistic changes, the whole structure cannot be decomposed because the other component does not occur systematically as a base in English (ex 5):

(5) *defend, repeal, collide, figment*

All these groups are marked as simple lexemes in this study, and their structure is not analysed further.

6.1.2.2. Complex lexemes

The rest of the sample was assigned to the category of complex lexemes and then their internal structure was examined further. The formal aspects that were studied are the type of base, i.e. free or bound, and the type of affix or word-formation process in terms of productivity. In addition, each complex lexeme was described as to the word-formation process involved. As far as the meaning is concerned, the lexemes were examined to find whether their meaning is compositional (i.e. transparent), and if not, whether they have only one, opaque sense, or whether they have both opaque and transparent senses. In addition, collocational anomalies were taken into account.

BNC SAMPLE COMPLEX LEXEMES	TYPE	TOTAL OF LEXEMES IN THE SAMPLE	TOTAL OF ITEMS ASSIGNED TO CATEGORIES
CATEGORY 0	regular		381
CATEGORY 1	formally anomalous		158
CATEGORY 2	collocationally anomalous		81
CATEGORY 3	semantically anomalous I		40
CATEGORY 4	semantically anomalous II		128
TOTAL		681	789

Table 6: Categories of complex lexemes in the BNC sample

On the basis of these aspects, the data were classified into categories 0-4 (see Table 6). Except for Category 0, all others are anomalous and therefore considered idiomatic. Since more anomalies often combine within one lexeme, each lexeme can be assigned to more than one category. As a result, there is a discrepancy between the number of complex lexemes in the

BNC sample (681) and the number of lexemes in the categories (789). The five categories are analysed and discussed below.

CATEGORY 0 is assigned to regular complex lexemes. These lexemes are semantically regular (i.e. transparent), formally regular (i.e. they contain a free base and are formed by productive processes), and also collocationally regular (there is no anomaly in the combination of constituents, i.e., there is no semantic incompatibility). The sample contains 381 complex lexemes of this type. Apart from derivatives and compounds, the complex lexemes in this category also include a group of combined formations, i.e. formations involving both composition and derivation. The distribution of the three types of complex lexemes in Category 0 in the sample is described in Table 7:

WF PROCESS	**BNC SAMPLE COMPLEX LEXEMES**	**%**	**CATEGORY 0 COMPLEX LEXEMES**	**%**
DERIVATIVE	523	76.8	311	81.6
COMPOUND	119	17.5	40	10.5
COMBINATION	39	5.7	30	7.9
TOTAL	681	100.0	381	100.0

Table 7: Distribution of word-formation processes in Category 0

A subgroup among the compounds in this category is formed by synthetic compounds (cf. Lieber, 2004) which are exemplified in (6):

(6) *record-breaking, government-owned, scriptwriter* and *screwdriver*

A similar group includes adjective compounds with an affixed head component:

(7) *bad-tempered, time-dependent, black-headed, air-conditioning, good-natured*.

Apart from these two groups, the category also includes neoclassical compounds extended by an affix (*physiologist, lexicographer*), and other lexemes whose formation involved two consecutive steps (*washing-machine, self-consciousness*).

These combined formations are for the most part transparent due to the high degree of descriptiveness: in terms of Štekauer's onomasiological theory, they represent the onomasiological type 1 (cf. 3.4.2.).

The table shows that the largest proportion of Category 0 lexemes is made up of derivatives, followed by compounds and combined formations. The proportion of each type corresponds roughly with the proportion of lexemes of the same type in the whole sample of complex words (with a slightly lower proportion of compounds in this category). This indicates that regular formations do not differ significantly in this respect from the whole class of complex words.

The following two tables illustrate the distribution of each word-class in Category 0:

WORD-CLASS	**BNC SAMPLE COMPLEX LEXEMES**	**%**	**CATEGORY 0 COMPLEX LEXEMES**	**%**
N	322	47.3	158	41.3
ADJ	264	38.8	171	45.0
ADV	47	6.9	40	10.5
V	48	7.0	12	3.2
TOTAL	681	100.0	381	100.0

Table 8: Word-class distribution in Category 0 and the BNC sample complex lexemes

WORD-CLASS	**BNC SAMPLE COMPLEX LEXEMES**	**CATEGORY 0 COMPLEX LEXEMES**	**BNC SAMPLE CL / CATEGORY 0 CL PROPORTION (%)**
N	322	158	48.8
ADJ	264	171	64.8
ADV	47	40	85.1
V	48	12	25.0
TOTAL	681	381	

Table 9: The word-class ratio between Category 0 and the BNC sample complex lexemes

As Table 8 illustrates, the distribution of word-classes in the complex lexemes of the BNC sample and in Category 0 of the sample is roughly similar, although there are differences: the adjectives and adverbs are slightly more frequent in Category 0 and, on the other hand, nouns and verbs are less frequent in this category.

The results shown by Table 9 are more interesting. The table shows the percentage of (complex) lexemes of each word-class in the BNC sample which are formed regularly (lexemes in Category 0), e.g. of the 47 adverbial complex lexemes in the BNC sample, 40 (85.1%) appear in Category 0, i.e. they are regular. By the same token, the results indicate that while the complex adverbs, but also adjectives tend to be formed regularly, the nouns (51.2%) and especially verbs (75.0%) display significantly more often some kind of anomaly. In fact, both these classes appear to be more anomalous than regular according to the sample data. We will first look more closely at the two groups of word-classes which are rather regular in formation and then come back to those which are irregular.

The group of adverbs is not very problematic. Derived adverbs are formed in English by quite a limited set of suffixes (*-ly, -wise, -ward(s), -ways, -s*). Of these affixes, *-ly* is highly productive. In fact, its productivity is so high that it is sometimes considered to be an inflectional morpheme in the literature (cf. Giegerich, 2012). However, other authors deny this, claiming that the suffix sometimes triggers semantic changes, which is uncommon for inflectional suffixes (e.g. *shortly, hardly, dryly*, cf. Plag 2003, 123), or they emphasize that the suffix functions in a system with other adverbial suffixes which are undoubtedly derivational (for more cf. Bauer, Lieber, & Plag, 2013: 323-237). It is no surprise then that all 40 instances of regular adverbs are formed by *–ly*, and that no other suffix appears in this category.

The group of adjectives is more heterogeneous, which follows from the fact that the range of adjectival suffixes is broader. However, there is also one group of derived adjectives which is very productive and bordering on inflection, i.e. participial adjectives. As Bauer, Lieber, & Plag (p. 306) explain, "[b]oth the *-ing* form of verbs and past participle forms are frequently used as premodifiers to nouns, and have sometimes been argued to be categorially adjectival, as evidenced by their frequent ability to accept prefixation with negative *un-*, to form the

comparative and superlative, sometimes even the morpohological comparative and superlative […] and to be sub-modified by *very, so,* etc." There are 62 non-idiomatic participial adjectives in the sample. However there are also 10 instances of semantically idiomatic participial compounds (e.g. *institutionalized, stilted, retiring, catching*), which shows that their status is sligtly different from the adverbial *-ly*.

In the following sections, we will proceed to the anomalous categories 1-4, starting with formal anomalies, proceeding to collocational anomalies (which are often linked to the anomalous form as well) and finaly we will describe two classes of semantically anomalous lexemes and discuss the combinations of several types of anomaly. The description thus proceeds from the periphery of the field of lexical idioms to its centre.

CATEGORY 1 contains words with a formal anomaly. The anomalies were divided into four groups: bound lexical bases, unproductive affixes, affixes of a different class and other anomalies.

Bound lexical bases are not used in productive word-formation processes, unless they are combining forms. Therefore, from the synchronic point of view they represent an anomaly, although they are certainly quite common in English. Bound lexical bases came into English with extensive borrowing from Romance languages and Greek and they are certainly peripheral with respect to the category of lexical idioms because they may be used systematically with Latinate affixes (cf. *nation, native, natality, neonate*). The category comprises 87 lexemes with bound bases in the sample, many of which combine this type of anomaly with one or more others. Here are some examples of lexemes with bound bases[4]:

(8) *granary, serial, renunciation, resumption, publicise, paralytic*

Unproductive affixes, like bound bases, do not participate in regular word-formation processes although they are very common in derivatives formed at some stage in history when they were productive (including both non-native and native affixes). In addition, a great many derivatives with unproductive affixes were borrowed into English as complete units, with

[4] It is to be noted that examples presented in this section often combine a formal anomaly with another one of the same or different type. They are to be understood as examples of the anomaly currently described and the other types are discussed elsewhere.

word-formation taking place in the source language. Example (9) illustrates the former type, (10) the latter type:

(9) *atonement, dilatory, reversal, breadth*

(10) *scripture, literate, scientific, pressure*

Another group which could be marked as formally anomalous is represented by lexemes which have a word-class specific suffix but belong to a different class. Two kinds of this phenomenon are illustrated below:

(11) *godly* (adj.), *disorderly* (adj.)

(12) *forward* (v.), *engineer* (v.), *pressure* (v.)

Example (11) illustrates the use of the adjectival suffix *–ly,* which may be considered anomalous in comparison with the almost fully productive adverbial *–ly*. However, the suffix is used systematically, and it is even moderately productive (cf. Bauer, Lieber & Plag, 2013: 306, who provide recent formations *demonly, dudely* and *speakerly*, and the web offers many similar facetious nonce words like "You bet your ass I acted *presidently*!"). It seems therefore plausible to consider its use as slightly anomalous, but such use of affixes is certainly not the only criterion of idiomaticity. A similar case is that of (ex 12) which represents lexemes with a different-class affix due to conversion. They are anomalous on account of signalling a different word-class by the affix, but at the same time the process of conversion is a common, productive process in English and therefore these instances are certainly not idiosyncratic.

All in all there are 64 lexemes with an unproductive affix, 126 lexemes contain a bound base and/or an unproductive affix.

The last group encompasses lexemes with formal anomalies of various kinds. There are 22 cases in this group. Some of the instances are exemplified below (ex 13) with their word-class and the specific anomaly described in brackets:

(13) *elderly* (ADJ, comparative base), *olden* (ADJ, archaic inflectional affix lexicalized), *topmost* (ADJ, anomalous superlative), *far-away* (ADJ, phrasal compound), *father-in-law* (N, phrasal compound), *stompie* (N, irregular base due to borrowing from Afrikaans), *showbiz* (N,

clipping compound), *casualty* (N, anomalous form of the affix *-ity*), *compstation* (N, clipping compound), *newfound* (ADJ+PP compound), *gunwale* (anomalous pronunciation of *wale*)

As can be seen, the group includes instances linked to inflection in one way or another (frozen inflectional affix, anomalous superlative, comparative base), examples with ties to syntax (phrasal compounds) and instances of irregular bases due to clipping and borrowing.

It is this last group of formally idiomatic lexemes in Category 1 that seems to be most relevant for the definition of idioms. It includes lexemes that one would probably call lexical idioms based on intuition only, in which they differ from words such as *nation, publicise* or *sweeten* from the former two categories. The plausibility of the criterial status of productivity is discussed below in 6.1.3.2.

The distribution of each word-class in Category 1 is illustrated by Table 10 below, while Table 11 shows the distribution of each word-formation process in this category:

WORD-CLASS	**BNC SAMPLE COMPLEX LEXEMES**	**%**	**CATEGORY 1 COMPLEX LEXEMES**	**%**
N	322	47.3	74	46.8
ADJ	264	38.8	56	35.4
ADV	47	6.9	3	1.9
V	48	7.0	25	15.9
TOTAL	681	100.0	158	100.0

Table 10: Word-class distribution in Category 1 and the BNC sample complex lexemes

WF PROCESS	BNC SAMPLE COMPLEX LEXEMES	%	CATEGORY 1 COMPLEX LEXEMES	%
DERIVATIVE	523	76.8	146	92.4
COMPOUND	119	17.5	12	7.6
COMBINATION	39	5.7	0	0
TOTAL	681	100.0	158	100.0

Table 11: Distribution of word-formation processes in Category 1 and the BNC sample complex lexemes

The distribution of word-classes (Table 10) differs significantly only for adverbs, which can be only expected due to the almost exclusive use of the productive *-ly* which is normally added to free bases (and therefore, these lexemes do not qualify for any of the first two subgroups) and for verbs, which, in contrast, occur more often within Category 1. The verbs in the category often have a bound base (*revolve, migrate, publicize, maximise*) and sometimes an unproductive affix (*pressure* (conversion)*, beware, encompass, sweeten*). There is also a formally idiomatic verb of the third group discussed above, *attune* (with an assimilation of the affix).

Neither is the distribution of word-formation types very surprising (Table 11). The vast majority of lexemes are derivatives due to the fact that both criteria of bound bases and unproductive affixes are (normally) limited to derivatives. The group of 12 compounds includes instances of idioms of the third type (anomalies of various kinds) and also a bunch of compounds with a monocollocable base (which is therefore bound) such as *mistletoe* and *Tuesday*. These lexemes are described in the following section.

CATEGORY 2 was assigned to lexemes with a collocational anomaly. The sample includes 81 instances assigned to this category. Collocational anomaly is generally of three types. The first type is represented by lexemes with a component which is either monocollocable (i.e. occurs only in the given combination and nowhere else; an example usually mentioned is *cran-* in *cranberry*) or which has very low collocability (occurs only in a very limited set of words, such as the often-quoted nominal affix *-th* in *length*). There are instances of both types in the sample. It is to be noted here that especially the category of lexemes with low collocability is very hard to define. To achieve some objective classification, each component

would have to be analysed in terms of its overall frequency of use and the frequency of occurrence within various lexemes, assuming that the threshold frequency of occurrences will be different for an affix (which is in the case of regular use expected to occur in a higher number of cases) and for a lexical base, especially of a specialized meaning (which could probably occur in a limited set of words even if regular in use). In addition, it is sometimes difficult to decide whether an anomaly is collocational or formal. The reason for this is that often the anomalous combination leads to an anomalous form.

The second type of anomaly can be described as formal incompatibility between the components. In this case, the complex word deviates from productive and regular word-formation processes by an unusual combination of components. Since this type of anomaly involves anomalous combinations of forms, it can be seen as a type on the borderline between collocational and formal anomaly.

The third type of anomaly is semantic incompatibility between the combined components. If the components are semantically incompatible, they do not make sense together in terms of their compositional meaning. It can be therefore pointed out that lexical idioms with semantically incompatible components are easier to be recognized as idioms for a hearer not acquainted with them. However, semantic incompatibility must be probably assigned by intuition only and many cases would be judged as compatible or incompatible by different evaluators depending on their imagination.

The first type described above, lexemes with a monocollocable component, includes a specific group which is associated by the origin and structure of lexemes with the first two types of formal anomaly (bound base, unproductive affix). I have dubbed this group as *non-native lexemes with an opaque base*, but the anomaly is linked rather to collocability in most cases. Lexemes in this category are of Romance or Greek origin borrowed as ready-made complex units. There are 38 examples in the category. The reason why these lexemes are treated separately is that they, similarly to the relevant groups of formal anomalies, have something in common. Below (ex 14) are some instances of the described type:

(14) *garrulous, duplicate, cadence, tirade, relegate*

Garrulous “loquacious” is formed from a bound base *garrul-*, which is however otherwise used only in *garrulity* (there are other lexemes, but these have the whole *garrulous-* as a base: *garrulously, garrulousness*). The meaning of the base can be therefore retrieved only from the meaning of the whole formation, and therefore, we can claim that within the whole formation, the base is opaque due to its low collocability. Similarly, the meaning of *duplic-, cad-, tir-,* or *releg-* cannot be retrieved based on analogy with a set of cognate English words.

A subgroup of this type is represented by lexemes whose bases occur in a set of relatively frequent words, but the meaning cannot be retrieved by reference to the word family because their meaning is too divergent. A good example of this type is the group *revolve, devolve, involve, evolve* and *convolve*. It is not easy to see any system in the use of *-volve* and its meaning in the complex words. Moreover, the same can be said for each of the words about the relation of the affix to the base *-volve*. Another example belonging to this category is the word family of *-pose*: *transpose, compose, expose, depose, dispose, prepose, propose*, etc. The meaning of the bound base *-pose* is not the same as the meaning of the verb *pose* and the meaning of the bound base, which could be generally described as “put to a position” is interpreted differently in each instance. These lexemes are certainly peripheral in the category of lexical idioms. They are close to the group mentioned in 1.1.2.1. (ex 5) in that they occupy the fuzzy area between simple and complex lexemes. In fact, Plag (2003: 32) describes similar examples as simple lexemes, claiming that “*infer, confer, prefer*, and *refer* are monomorphemic words, because there are no meaningful units discernible that are smaller than the whole word”. They can be seen in the continuum of this category in the direction of complex lexemes, nevertheless they still display some qualities of simple words.

Apart from this group of opaque formations, there is a group of other monocollocable formations (or formations with low collocability). They still belong to the first type of collocational anomaly and are central in the category of idioms because their anomalies are really idiosyncratic (ex 15):

(15) *mohair, Tuesday, mulberry*, *mistletoe, best, registrar, carpenter*

Mohair is originally a loanword in which the second part was identified with the form *hair* by folk etymology. This has led to *mo-* becoming a monocollocable component with opaque

meaning. *Tuesday* consists of a common lexical morpheme *day* and a monocollocable *tues-* which has developed into this opaque form from the regular genitive case of *Tiw*. *Mulberry* is similar to the often-mentioned *cranberry*; *mistle-* in *mistletoe* occurs also alone, but normally only in the collocation *mistle thrush*. *Be-* in the superlative *best* is the suppletive form of *good* and it is a borderline case between inflection and derivation. As for *registrar*, suffix *-ar* as a nominal agentive suffix is marginal in comparison with *-er* and *-or*. *Carpenter* has the form of an agentive noun, but the verb *carpent* is rare, and the combination is therefore anomalous in comparison with regular pairs such as *act – actor, speak – speaker*, etc.

Monocollocable components can also come in the form of clipped bases in clipping compounds such as *compstation* or *showbizz*.

The second type, formal incompatibility between the components, is represented by lexemes resulting from a formally anomalous combination of components. Some instances are an idiosyncratic anomaly, some are members of a less prototypical word-formation pattern (ex 16):

(16) *dogged, walling, urinal, polyunsaturated, consumerism*

The suffixes *-ing* and *-ed* are prototypically used in participles and deverbal adjectives and nouns. However, in *walling* and *dogged,* they are added to nouns. In some cases, such as *campaigning* or *banded*, it is difficult to decide whether the adjective is formed from a noun or a verb. These words are considered regular (as if deverbal) in the sample. In *urinal*, *-al* is added to a noun to form another noun, which is anomalous. *Polyunsaturated* contains two formally and semantically incompatible prefixes (this word is a term, for the discussion of terms, see 6.1.3.6.) and *consumerism* contains an agentive suffix before *-ism*.

The third type, semantic incompatibility, was recognized mostly in compounds (20 compounds, 4 derivatives and 1 combined formation). This is not surprising given the fact that incompatibility of two lexical meanings is more probable due to a large number of lexical fields and specificity of lexical meanings, than semantic incompatibility between a lexical component and a grammatical component (which is more general in meaning and thus more compatible). Instances of semantic incompatibility are illustrated by (ex 17):

(17) *deadlock, spendthrift, rainbow, goalmouth, starfish, wholesale, interface*

Deadlock and *spendthrift* are prototypical examples of lexical idioms. The semantic incompatibility between their components is very strong: a *lock* can have no such quality as being *dead,* the usual meaning of *thrift* is in direct opposition to *spend.* In *deadlock*, the first component is used in the metaphorical sense "absolute", whereas in the latter lexeme, *thrift* is used in the meaning "economical management" (which however is otherwise not a common meaning of *thrift*). *Rain* and *bow* are incompatible (a rain cannot form or have a bow) and *rainbow* is based on metonymy (the association with rain), a *goal* has no *mouth* (*goalmouth* is based on metaphor), *star* in *starfish* has nothing to do with *fish* (except when used in metaphorical reference to shape) and it is unclear how a *sale* can be *whole* in *wholesale*. The last example, *interface*, is a derivative in which it is difficult to see how exactly *inter-* and *-face* can be related.

Distribution of the Category 2 complex lexemes in terms of their word-class in the BNC sample is clear from Table 12. Table 13 illustrates the distribution of each word-formation process this category:

WORD-CLASS	**BNC SAMPLE COMPLEX LEXEMES**	**%**	**CATEGORY 2 COMPLEX LEXEMES**	**%**
N	322	47.3	49	60.5
ADJ	264	38.8	15	18.5
ADV	47	6.9	3	3.7
V	48	7.0	14	17.3
TOTAL	681	100.0	81	100.0

Table 12: Word-class distribution in Category 2 and the BNC sample complex lexemes

WF PROCESS	BNC SAMPLE COMPLEX LEXEMES	%	CATEGORY 2 COMPLEX LEXEMES	%
DERIVATIVE	523	76.8	51	63
COMPOUND	119	17.5	29	35.8
COMBINATION	39	5.7	1	1.2
TOTAL	681	100.0	81	100.0

Table 13: Distribution of word-formation processes in Category 2 and the BNC sample complex lexemes

The results indicate a higher number of nouns and a higher number of compounds among the Category 2 lexemes than in the whole sample of complex words. This reflects the fact that collocational anomalies include semantic aspects, which appear to be more common in nominal compounds. The higher number of verbs is due to Latinate borrowings with bound and opaque bases (*prostrate, vaccilate, placate,* etc.).

CATEGORY 3 is the first type of semantic anomaly. Lexemes included in this group are less idiomatic than lexemes included in Category 4 below. The meaning of a lexeme of this category is non-compositional in one or more of its senses. The non-compositionality is however based on some relatively transparent semantic shift: metaphor, metonymy, meaning specialization or generalization, and the word-formation sense is retained side by side with the shifted sense, or the shifted sense is relatively close to the word-formation sense (especially in meaning specialization and generalization). Because of this partial transparency, these lexemes could probably be considered non-idiomatic in some classifications. The peripheral position in the category of lexical idioms is also supported by the finding that lexemes of this category in the sample almost never display another type of anomaly. It is probably the case, however, that when the shifted meaning is used conventionally (i.e. lexicalized), it becomes partly independent of the original non-idiomatic meaning. The speaker knows that this specific non-compositional meaning is linked to this lexeme, and therefore, it is idiomatic in

this non-compositional sense. There are 41 lexemes assigned to Category 3 in the sample and some of them are exemplified by (18) to (20):

(18) *mouth-watering, tubby, offshoot, goer, unearth*

(19) *craftsmanship, oddity, skinny*

(20) *protector, tipper*

Examples in (18) represent idioms based on metaphor. All of them also retain their non-idiomatic sense. Most of them are central examples with clear membership in this category:

Mouth-watering is non-idiomatic when used with food (although it may also be considered slightly idiomatic in that it may be hard to distinguish the fully non-idiomatic use, i.e. "something that really causes salivation", from the metaphoric description of something which looks delicious, cf. Moon's discussion of *shake hands* or *raise one's eyebrows* in 2.5.4.). However, *mouth-watering* can also be used with something attractive, but non-edible (BNC gives examples of use its with *mountain, image, ideas, plans*, etc.) and in this case the expression is truly metaphoric. *Tubby* is used about something "shaped like a tub" (non-idiomatic) or about someone "plump"[5] (idiomatic). *Offshoot* is "a shoot or branch growing from the main stem" (*shoot* is metaphorical, but the shift precedes the composition) or "something that develops or derives from a principal source or origin" (idiomatic). *Goer* is slightly more problematic as the meaning of both components is very vague. However, we can still consider the meaning "a person who attends something regularly" as non-idiomatic to only weakly idiomatic, whereas the meaning "an energetic person" or "an acceptable or feasible idea, proposal, etc." are definitely idiomatic. *Unearth* is an example of a verbal member of the category, meaning "to dig up out of the earth" in the non-idiomatic sense and "to reveal or discover" in the idiomatic one.

Examples in (19) represent idioms based on metonymy: *craftsmanship* can denote either "the skill that someone uses when they make beautiful things with their hands" (non-idiomatic) or "the quality that something has when it is beautiful and has been very carefully made" (idiomatic). *Oddity* as "an odd quality or characteristic" is non-idiomatic, but "an odd person

[5] Note that only selected senses are described here for simplification.

or thing" is idiomatic. *Skinny* as "consisting of or resembling skin" is non-idiomatic and "lacking in flesh; thin" is idiomatic.

Examples in (20) are instances of meaning specialization: *Protector* is either "a person or thing that protects" (compositional) or "a person who exercised royal authority during the minority, absence, or incapacity of the monarch" (meaning specialization). *Tipper* as "a kind of truck which can be tipped" is meaning specialization, whereas "someone who makes something tip" or "something that tips" is the general, compositional meaning (*tipper* as a "person who gives tips" is semantically transparent but formally anomalous).

The following tables (Table 14 and 15) summarize the distribution of Category 3 lexemes. The distribution of lexemes in terms of word-classes and word-formation process does not differ significantly from the whole complex lexemes sample. Moreover, the number of lexemes is rather low to provide relevant data.

WORD-CLASS	**BNC SAMPLE COMPLEX LEXEMES**	**%**	**CATEGORY 3 COMPLEX LEXEMES**	**%**
N	322	47.3	21	51.2
ADJ	264	38.8	17	41.5
ADV	47	6.9	0	0
V	48	7.0	3	7.3
TOTAL	681	100.0	41	100.0

Table 14: Word-class distribution in Category 3 and the BNC sample complex lexemes

WF PROCESS	**BNC SAMPLE COMPLEX LEXEMES**	**%**	**CATEGORY 3 COMPLEX LEXEMES**	**%**
DERIVATIVE	523	76.8	30	73.1
COMPOUND	119	17.5	9	22.0
COMBINATION	39	5.7	2	4.9
TOTAL	681	100.0	41	100.0

Table 15: Distribution of word-formation processes in Category 3 and the BNC sample complex lexemes

CATEGORY 4 includes lexemes which are also semantically anomalous. Unlike Category 3 lexemes, however, Category 4 lexemes do not have both word-formation meaning and shifted idiomatic meaning. Their meaning is either shifted (by metaphor, metonymy, etc.) completely or the semantic anomaly is of a different, often opaque, type. There are 128 instances in the sample. The following part describes some of the semantic types in Category 4 (exx 21-26):

(21) *pin-point, catching, foreman*

(22) *mindful, flipper, on-board*

(23) *institutionalized, couplet, washing-machine*

(24) *opportunism*

(25) *collegiality*

(26) *foolproof*

Examples in (21) are instances of metaphor. It is interesting that metaphor-based lexemes are not very frequent in this group. *Pin-point* "to locate or identify exactly" is based upon similarity between a precise point and the sharp end of a pin. *Catching* "infectious" or "captivating" is a metaphoric concrete-to-abstract shift. *Foreman* "a person, often experienced, who supervises other workmen" stands at the front only metaphorically.

Examples in (22) are based on metonymy: *mindful* "keeping aware" is based upon association between consciousness and mind, *flipper* "the flat broad limb of seals, whales, penguins, etc." is based on association with its movement (it could also be seen as meaning specialization or "something that flips") and *on-board* meaning "on or in a ship, boat, aeroplane, or other vehicle", is an instance of synecdoche.

The largest number of examples based on a semantic shift are based on meaning specialization, which is common in the process of lexicalization. Meaning specialization is exemplified in (23): *institutionalized* "placed in an institution, esp. a psychiatric hospital or penal institution or a children's home or home for elderly people" is more specific than its word-formation meaning, as well as *washing-machine* "a machine for washing clothes" and

couplet “two successive lines of verse”. The latter example can be also classified as a literary term. Terms are often based on meaning specialization (for more see later)

There are also occasional instances of meaning deterioration (ex 24), amelioration (ex 25) and pejoration (ex 26).

Other, non-systematic kinds of semantic anomaly include (ex 27):

(27) *double-sided, nursery, grand-daughter, shorthand*

A *double-sided* thing is “usable on both sides”, *nursery* is “a room in a house set apart for use by children”, *grand-* in *grand-daughter* has little to do with the common use of the adjective *grand*, and *shorthand* “a system of speed writing” is even more idiomatic by describing something which is neither *short*, nor *hand*. The latter example is also an example of an exocentric compound. Other examples of this type of lexical idioms are in (28):

(28) *dreadnought, pullover, spendthrift*

Another formally specified type is the particle compound exemplified in (29):

(29) *logon, roll-out, break-out, overhaul*

From one point of view, particle compounds, i.e. compound nouns formed from phrasal verbs, are indisputably lexical idioms in one aspect: they are analogous to phrasal verbs which are generally counted among idioms because of their polylexicality, opaque meaning and fixed structure. However, they are also in most cases formed regularly from the corresponding phrasal verbs and their meaning corresponds to the meaning of the phrasal verbs. This problem is also associated with a more general question of how to decompose complex units when their transparency or opaqueness is assessed (see section 6.1.3.5.)

The distribution of word-classes and word-formation processes in Category 4 within sample is presented in Tables 16 and 17.

WORD-CLASS	BNC SAMPLE COMPLEX LEXEMES	%	CATEGORY 4 COMPLEX LEXEMES	%
N	322	47.3	86	67.2
ADJ	264	38.8	26	20.3
ADV	47	6.9	4	3.1
V	48	7.0	12	9.4
TOTAL	681	100.0	128	100.0

Table 16: Word-class distribution in Category 4 and the BNC sample complex lexemes

WF PROCESS	BNC SAMPLE COMPLEX LEXEMES	%	CATEGORY 4 COMPLEX LEXEMES	%
DERIVATIVE	523	76.8	60	46.9
COMPOUND	119	17.5	61	47.6
COMBINATION	39	5.7	7	5.5
TOTAL	681	100	128	100

Table 17: Distribution of word-formation processes in Category 4 and the BNC sample complex lexemes

The tables indicate that semantic anomaly is definitely more common for some types of lexemes. Table 16 shows a considerably higher occurrence of semantically anomalous nouns and slightly higher numbers are also seen for verbs. Table 17 suggests that compounds incline to anomalous meaning (which is also in accordance with the expectations based on a high number of mentions of idiomatic compounds in literature).

6.1.3. Definition of lexical idioms revisited

The above analysis of a sample of 1 000 random lexemes from the BNC has shown that when analysing English vocabulary in terms of phraseology, one has to deal with various degrees and subtypes of each of the three main types of anomaly. Moreover, sometimes it is even problematic to assign a lexeme unequivocally to the category of simple or complex lexemes. The following section presents a summary of the problematic areas which need to be dealt with when analysing English lexical idioms.

6.1.3.1. Simple vs. complex lexemes

Apart from clearly simple and clearly complex lexemes, English vocabulary also contains a significant number of words which are only partly analysable in terms of structure, but not fully. This concerns especially non-native words with an affix not unfamiliar to the speaker of English. The affix can occur in more than one formation, but the base (if it can be called a base in lexemes borrowed as a complete unit) is not productively used in English. Although it would be possible to regard all these borrowed lexemes as simple lexemes in the English word-formation system, I have drawn a simple/complex borderline slightly further within the field, keeping in the category of simple lexemes also lexemes which have a formally distinct, but semantically rather opaque, affix and a base which is not used in English systematically (*collide, defend*). On the other hand, complex lexemes in this study include instances with a bound base and a distinct affix (both semantically and formally) even if the base is used in a very limited set of lexemes or is opaque due to lexicalization (*garrulous, transpose, include*). It is nevertheless evident that this borderline between simple and complex is somewhat arbitrary and that the lexemes on either side of and close to the borderline all exhibit the same quality, i.e. the degree of lexicalization, influenced in addition by the randomness of linguistic borrowing and other factors.

6.1.3.2. The criterion of productivity

One of the main questions arising in connection with the defining criteria of lexical idioms is whether the criterion of (non-)productivity should be included in the list of relevant formal anomalies of lexical idioms. The main reason given as to why it should is that speakers must store and retrieve a formation based upon unproductive processes as one unit, without segmenting it into its components. However, there is no real evidence that this condition really distinguishes unproductively formed words from those formed productively. The assumption that regular combinations are stored and retrieved by segments while irregular combinations are stored and retrieved as whole units is much less problematic on the syntactic level, where the productivity rules generally have much greater validity. Still recent psycholinguistic research indicates that even compositional reoccurring units such as lexical bundles (cf. Biber et al., 1999) differ from free combinations of words in the way they are processed (cf. Tremblay et al., 2011). As far as complex lexemes are concerned, Aitchison

(2012: 152-155) supports the idea of words being stored as one unit, arguing among others by pointing out the mechanism of slips of the tongue (where it is not usually the suffix which is used wrongly). She concludes: "On the whole, experiments have shown three facts: first, findings on English do not necessarily generalize to other languages, or vice versa. Second, in English, it does not take any longer to recognize a word with a derivational suffix […] than a word without. […] A third finding is that people can split words up if they need to" (p. 154).

In my opinion, unproductivity should not be considered a defining criterion of lexical idioms in English because it does not represent an idiosyncratic phenomenon for two main reasons. First, the amount of unproductively formed English words is quite high. In fact, 18.6 % of complex words in the sample (126 out of 681) are formed in this manner, which indicates that they are not exceptions in the proper sense of the word. Indeed, an overwhelming majority of these words are of common origin (Latin or Greek loanwords and words formed within English by analogy with these Latin and Greek words) and therefore, they are not idiosyncratic instances of irregularity – they exist rather within a system with certain rules, although the rules are not used (normally) anymore to form new lexemes. Second, as the psycholinguistic research shows, we cannot approach the production of words in the same way as we approach the production of word combinations. The form of an actual word is always fixed and the free choice of a suffix is only theoretical because of such factors as lexical blocking (i.e. non-occurrence of a word-form, whose existence could be expected based on productive rules, due to the existence of a rival form, cf. Plag, 2003). In view of the generally high occurrence of formal anomalies at the level of a word, I therefore assume that the overall relevance of formal anomalies in both production and (especially) perception is smaller than that of other types of anomaly and that unproductivity should better be excluded from the defining criteria of English phraseological units.

6.1.3.3. The priority of semantic criterion

In the previous section I have argued that the formal criterion is problematic due to lexicalization (i.e. complex units functioning as lexemes are formally fixed by default) and the high overall occurrence rate of words formed by synchronically unproductive processes. The collocational criterion plays its role as far as formal incompatibility is concerned, but this

happens only marginally. Again, there is a large group of opaque and unsystematically used Latinate (and Greek) bases. The formations using them are however on the borderline with simple lexemes, not with idiomatic lexemes. Therefore, semantics remains the most important aspect of idiomatic expressions: both the anomaly of the components (here referred to as *semantic anomaly*) and the anomaly of their combinations (here referred to as *collocational semantic anomaly*). Collocational semantic anomaly occurs typically together with semantic anomaly of components (because semantically incompatible components can be combined into a meaningful unit only if their meaning is different from the regular meaning).

6.1.3.4. Lexemes formed within English vs. lexemes formed outside English

The previous questions lead to another, more general question: Should phraseology consider all complex lexemes in vocabulary in an attempt to identify lexical idioms, or should it only concern itself with lexemes formed within English? This is a question which is not discussed in the phraseological theories because the problem is specific to single-word units[6]. I do not claim to have a definite answer to this question, but it may be useful to look at it more closely. In my opinion, the answer depends on the aim of the phraseological research. In a structuralist theoretical approach, one of the aims may be to identify all complex units anomalous with respect to the currently valid rules. This seems to be the approach of Čermák (2007a: 76) when he writes that "[b]y **regular combinations** in a generally syntagmatic sense we mean all combinations governed by analogous rules, i.e. rules based on analogy, […] Thus PI goes a step further in the area of combinatory outcomes of the extensions of combinatory possibilities in language, crossing the boundaries and operating in an area of combinations that, according to the standard rules of language, cannot or should not take place, i.e. **anomalous combinations**." As far as formerly productive combinations are concerned, he adds: "Diachronically anomalous phraseological combinations came into being in various ways, in collocational phrasemes especially through a simile and individual metaphors, and may, among other things, reflect the residuum of combinatory possibilities previously common." This section thus seems to refer to combinations formed only within the studied language. However, it seems that English is because of the specific structure of its vocabulary

[6] Borrowing is sometimes discussed in connection to phraseology but only in the sense of loan translations of multi-word units (cf. Gottlieb 2012).

different from Czech and there is a clash between the system of derivatives (or compounds) and the system of word-formation processes productive at some stage in English. It should be also added that standard accounts of word-formation (e.g. Huddleston & Pullum, 2002, or Bauer, Lieber & Plag, 2013) describe both words borrowed as one unit and words formed within English. It seems therefore that it is first of all necessary to define the aim, i.e. do we want to study existing combinations with fixed structure created by currently unproductive processes to see the proportion of fixed and free combinations in vocabulary? Or do we want to focus on idiosyncrasies of an unexpected character, wishing for example to use the results in ELT or other applied disciplines? Or, finally, do we want to study the system from the theoretical point of view but focus on the existing system of analogous rules rather than the current word-formation rules?

In my opinion, the most plausible scope of phraseological research for English is such that it excludes, besides simple words and opaque complex borrowings, also complex words which are formed by non-productive rules but still used in English to such an extent that the speakers may be expected to associate their components with some grammatical or semantic meaning. Using this elimination rule, words such as *insurance, guidance, resistance* or *closure, composure, departure,* which are derivatives with an unproductive affix, will be excluded from the category of idioms. However, although productivity will be rejected as a defining criterion of lexical idioms, these lexemes can still be included in this category provided they contain another type of anomaly (e.g. *allotment*, "a small area of land in a town which a person can rent in order to grow fruit and vegetables on it"). Also, under this approach, borrowings with distinct components (at least partly associated with recognizable meaning) can also be included, as from the synchronic point of view they do not differ from analogous English formations (e.g. *surmount*).

6.1.3.5. Decomposition for semantic analysis

Another theoretical stumbling block which emerged in the process of practical analysis is the problem of deciding which components should be assessed when studying semantic anomaly of words composed of three or more morphemes. Examples from the BNC sample include:

(30) *dismemberment*, *craftsmanship*, *shamefaced*, *set-back, unrecognizable*

From the formal point of view, *dismemberment* is formed regularly (although *-ment* is only marginally productive at present): *dis-* is a transpositional prefix changing a noun to a verb, and *-ment* is a deverbal nominal suffix. It is evident that the word has been formed in these two stages: [[dis-member]-ment]. However, how should the word be analysed in terms of semantics? Should we regard *dismemberment* as transparent, because the meaning of *dismember* is analogous to the meaning of *dismemberment*? From the structurally-theoretical point of view, this would probably be the correct analysis. However, for some purposes (e.g. language acquisition, ELT) it would make more sense to treat the lexeme as idiomatic, because although the (three) basic components are all common in English, their sum does not correspond to the overall meaning (we could add that *dismember* is less likely to be known to the speaker than its individual parts).

The second example, *craftsmanship*, also consists of several components: [[[craft-s]-man]-ship]. Intuitively, it does not seem useful to go beyond the level of *craftsman*. This may be because there is nothing semantically anomalous below the level, but there may be other reasons for this, for example *craftsmanship* differs from *dismemberment* in the fact that *craftsman* is not considerably different in frequency from *craftsmanship.*

Shamefaced is an instance of a synthetic adjective in which both word-formation processes, composition and derivation, are usually thought to have taken place at one stage: [shame-face-ed]. In this case, there is therefore no other possibility than to analyse all three parts.

Set-back is an instance of an arbitrary particle compound (there are many similar instances in the sample) which does not consist of three morphemes but is formed from a complex base. As a typical instance, the nominal meaning of the compound is analogous to the verbal meaning. Therefore, the stance we take to the theoretical question of decomposition preferences directly influences whether the sample of lexical idioms will include such particle compounds or not.

As far as the last example, *recognizable,* is concerned, it seems much better to decompose it into only three parts [*un*-[*recognize-able*]]. This intuitive decision is probably due to the unproductively formed base of the second step – one tends to see it as one unit.

These examples suggest that there are probably several factors which influence our perception of complex words. In the present analysis the lexemes were decomposed into its constituent elements if the structure of the base was not too frozen and opaque due to diachronic changes. Nevertheless, the issue has not been studied in detail and the current approach is partially based on intuition.

6.1.3.6. Terms in the sample

The BNC sample contains also a number of terms (for the definition of terms, cf. 2.5.5.4.). There are 36 terms which were not assigned to any of the anomalous categories (they are assigned to Category 0) because they cannot be analysed in terms of phraseological criteria. They include: *creatinine, bromide* and similar concepts from natural sciences, but also *incrementalism* and *nihilism* from humanities. Apart from these central terms, there are also several quasiphrasemes from folk terminology such as *rattlesnake, blackbird* and other lexemes on the borderline between terminology and general language (*pluralist, prepatent, free-kick*). The quasiphrasemes have been assigned to a category according to their anomalies, but they are considered marginal in this study.

6.1.3.7. The new definition of lexical idioms

The provisional definition (cf. 5.3.) was formulated as follows:

The lexical idiom is a single-word lexeme formed as a combination of components which is anomalous semantically and/or collocationally and/or grammatically.

It can now be amended on the basis of the above findings:

The lexical idiom is a single-word lexeme formed as a combination of components which is semantically anomalous and in addition can also exhibit a formal or collocational anomaly.

It must be also added that:

By anomaly we mean all deviations from the systematic use of the same or analogous components within English, regardless of whether the word-formation process is currently productive or not.

The occurrence of opaque Latinate or Greek bases due to borrowing and/or a high degree of lexicalization is such a frequent phenomenon in English that these instances are better considered borderline cases close to simple lexemes rather than instances of word-formation with anomalous rules.

The sample has also shown that due to the specific status of words as units fixed by lexicalization (cf. Klötzerová, 1997, see also 4.4.), the main type of formal anomaly, formal fixedness, cannot be used as a criterion of idiomaticity as formal fixedness is common to all institutionalized words. In addition, since it seems to be reasonable in English not to include the criterion the productivity (cf. 6.1.3.2.), the range of anomalies which can be marked as formal has basically shrank to anomalous combinations of components, which is, however, a type of anomaly on the borderline between the formal and semantic anomaly. In the following analysis, we will treat this type of anomaly as within the category of formal anomalies to distinguish it from an anomalous combination of meanings (semantic incompatibility).

It is to be emphasized that the various types of anomalies are scalar in nature and that those resulting in a lower degree of idiomaticity (e.g. collocational anomaly of the type N + *-ing* in *walling*) lead to the peripheral position of these lexemes within the category or contribute to the overall idiomaticity when the lower-degree anomaly is combined with another one (e.g. N + *-ed* adjective *dogged* "showing determination; not giving up easily").

6.1.4. Central lexical idioms in the BNC sample

By applying these rules, we can also retrieve a group of lexemes from the sample that occupy the most central position among lexical idioms. These are lexemes combining the semantic anomaly (Categories 1 or 2) with a relevant type of formal or collocational anomaly. All lexemes which meet these criteria are listed in Table 18 and Tables 19-22 illustrate various aspects of this category:

lexeme	word-class	semantic anomaly	formal anomaly	collocational anomaly – form	collocational anomaly – meaning
biplane	N	**X**			**X**
bloodstock	N	**X**			**X**
casualty	N	**X**	**X**		
compstation	N	**X**	**X**	**X**	
consumerism	N	**X**	**X**	**X**	
data-base	N	**X**			**X**
deadlock	N	**X**			**X**
defunct	ADJ	**X**	**X**	**X**	
dismemberment	N	**X**			**X**
dogged	ADJ	**X**		**X**	**X**
engineer	VERB	**X**	**X**		
father-in-law	N	**X**	**X**	**X**	
forward	VERB	**X**	**X**		
goalkeeping	N	**X**			**X**
goalmouth	N	**X**			**X**
godly	ADJ	**X**	**X**		
gunwale	SUBS	**X**	**X**		
hereby	ADV	**X**		**X**	**X**
check-in	N	**X**			**X**
interface	VERB	**X**			**X**
livestock	N	**X**			**X**
mistletoe	N	**X**	**X**	**X**	
mulberry	N	**X**	**X**	**X**	
onset	N	**X**			**X**
overhaul	VERB	**X**			**X**
payroll	N	**X**			**X**
rainbow	N	**X**			**X**
seascape	N	**X**			**X**
showbiz	N	**X**	**X**	**X**	
spendthrift	N	**X**			**X**
starfish	N	**X**			**X**
stompie	N	**X**	**X**		
underhand	ADJ	**X**	**X**		
understudy	N	**X**			**X**
underway	ADV	**X**	**X**		
upkeep	N	**X**			**X**
urinal	N	**X**		**X**	
wholesale	ADJ	**X**			**X**
workload	N	**X**			**X**

Table 18: Idiomatic lexemes with combined anomalies in the BNC sample

WORD-CLASS	COMPLEX LEXEMES	%	CENTRAL LEXICAL IDIOMS	%
N	322	47.3	28	71.8
ADJ	264	38.8	5	12.8
ADV	47	6.9	2	5.1
V	48	7.0	4	10.3
TOTAL	681	100.0	39	100.0

Table 19: Word-class distribution among central lexical idioms

WF PROCESS	COMPLEX LEXEMES	%	CENTRAL LEXICAL IDIOMS	%
DERIVATIVE	523	76.8	13	33.3
COMPOUND	119	17.5	25	64.1
COMBINATION	39	5.7	1	2.6
TOTAL	681	100.0	39	100.0

Table 20: Distribution of word-formation processes among central lexical idioms

WORD-CLASS	REGULAR COMPLEX LEXEMES AVERAGE FREQUENCY	CENTRAL LEXICAL IDIOMS AVERAGE FREQUENCY
N	789	260
ADJ	478	220
ADV	452	420
V	180	264

Table 21: Average frequency in BNC of regular and idiomatic lexemes with respect to word-classes.

WF PROCESS	REGULAR COMPLEX LEXEMES AVERAGE FREQUENCY	CENTRAL LEXICAL IDIOMS AVERAGE FREQUENCY
DERIVATIVE	670	255
COMPOUND	261	275
COMBINATION	83	77

Table 22: Average frequency in BNC of regular and idiomatic lexemes with respect to word-formation processes

Although the sample of central lexical idioms is quite small to provide valid data, we can certainly see at least some tendencies in comparison with the rest of the sample. First, the proportion of nouns within this group is much higher than in the whole sample of complex words (Table 19). The same can be said about compounds as far as word-formation is concerned (Table 20). On the other hand, other word-classes are less numerous and the same applies to derivatives. Combined formations are only marginal due to their complex onomasiological structure which contains all three basic components, i.e. the base, the determining and the determined constituents of the mark.

There are also interesting differences in the average occurrence of the different groups of lexemes in the BNC sample. This time it is useful to compare the group of central idioms with the lexemes of Category 0 (regular complex formations). This comparison may show whether idiomatic compounds tend to be more or less frequent than regular formations. Tables 21 and 22 suggest that regular nouns and adjectives / derivatives are more frequent than corresponding idiomatic units. However, the results are biased by a few very frequent lexemes in Category 0 which considerably raise the average occurrence (the median value is 86 for Category 0 and 117 for lexical idioms). This may imply that lexical idioms are slightly more common, but a further analysis based on a larger sample is be needed.

To conclude the analysis of the first sample, the main outcome of the analysis is that its findings made it possible to arrive at an operational set of criteria for identifying lexical idioms: the vast majority of lexemes presented in Table 18 are of the kind speakers would intuitively call lexical idioms. They are of course only the core of the lexemes that show a combination of anomalies of different kinds. Apart from these, there are other instances of semantically strongly idiomatic lexemes in the sample which however do not combine semantic anomaly with another one (they includes e.g. *interview, nursery* and most particle compounds such as *uptake*).

6.2. The OED sample analysis

6.2.1. Collection and analysis of the OED sample

The general properties of the basis for the OED sample are described in 5.5. In summary, the OED sample contains:

- lexemes with the first entry into OED between years 1800 and 2017
- complex lexemes
- lexemes formed within English (not loanwords)

The main criterion for inclusion of a lexeme into the sample is semantic anomaly, i.e. non-compositionality in at least one of the lexeme's relevant senses. There certainly needs to be a borderline to exclude really marginal senses of lexemes (e.g. specialized vocabulary and rare senses). Therefore, I consider those senses relevant which are included in *Collins English Dictionary online*[7] (hereafter referred to as CD), which excludes the most marginal senses listed in OED, comparing sometimes data with data from *The Oxford Advanced Learner's Dictionary*[8] (hereafter referred to as OALD).

The chosen 500 lexemes with semantic anomaly are then studied further: the semantic anomaly is specified in greater detail to distinguish systematic meaning shifts from cases of more opaque semantic structure. The aim of the analysis in this respect is to find and describe both systematic and idiosyncratic instances of non-compositionality and possibly make a decision about their position within the category of lexical idioms.

The lexemes are then analysed in terms of formal and collocational anomaly, working on the assumption that a second (or third) type of anomaly will increase their overall degree of idiomaticity. Semantic anomaly is also assigned to lexemes with an opaque, monocollocable component or an idiosyncratic formal structure. This is because all these anomalies are reflected secondarily in the semantic non-compositionality of a lexeme.

[7] https://www.collinsdictionary.com
[8] https://www.oxfordlearnersdictionaries.com

In addition, several other factors are studied: the word-class of the lexeme, its word-formation type, and two features associated with its use: membership in a specific field of vocabulary and pragmatic function.

Since entries in OED usually subsume closely related word-classes (e.g. *far-away*, adj., adv., and n.) and the idiomatic features are usually different for each word-class (e.g. formal features, but also semantic features due to semantic shift), only one word-class was chosen to be included in the sample for simplification. Where the degree of idiomaticity is similar, the primary, basic, word-class was chosen (i.e. adverb for *far-away*). However, if the secondary use presents a new, unexpected type of anomaly, the secondary word-class is chosen (e.g. for *dugout,* adj. and n., the nominal meaning "a small boat that is made by removing the inside of a log" or "a shelter made by digging a hole in the ground" is chosen instead of the primary adjectival meaning derived directly from the phrasal verb). This approach is taken in order to introduce a wider range of idiomatic types into the sample.

6.2.2. Description of the sample

The sample contains 500 lexemes. The number of complex lexemes formed within English needed to make up this sample was 1 662. In addition, approximately 20% of the whole set of complex lexemes were terms, which were not included in the sample (unless bordering on general language, cf. 6.2.6) because of their specific referential status within vocabulary. If we combine these data, we will come to the conclusion that about 38% of non-terminological vocabulary are anomalous in meaning. However, the further analysis will show that some subgroups of non-compositional lexemes are still very marginal within the category of lexical idioms and that the number corresponds to lexical idioms in a very broad sense.

As far as the word-formation type is concerned, the study does not distinguish only compounds and derivatives (similarly to Čermák (2007a)), but also adds the category of combined formations, which includes lexemes with a combination of both composition and derivation, either as two subsequent steps (e.g. *broadcasting*) or as one word-formation step including two simultaneous processes as in synthetic compound nouns (e.g. *care-taker*) or

some adjectives (e.g. *one-dimensional*). The category of other formations includes atypical formations such as blends and clipping formations.

Tables 23 and 24 illustrate the distribution of word-classes and word-formation processes in the OED sample and Figure 8 illustrates the relation of these two formal aspects:

WORD-CLASS	**OED COMPLEX LEXEMES**	
	NUMBER	**%**
N	368	73.6
ADJ	97	19.4
V	30	6.0
ADV	5	1.0
TOTAL	500	100.0

Table 23: Distribution of word-classes in the OED sample

WF PROCESS	**OED COMPLEX LEXEMES**	
	NUMBER	**%**
COMPOUNDS	283	56.6
DERIVATIVES	158	31.6
COMBINED FORMATIONS	44	8.8
OTHER	15	3.0
TOTAL	500	100.0

Table 24: Distribution of word-formation processes in the OED sample

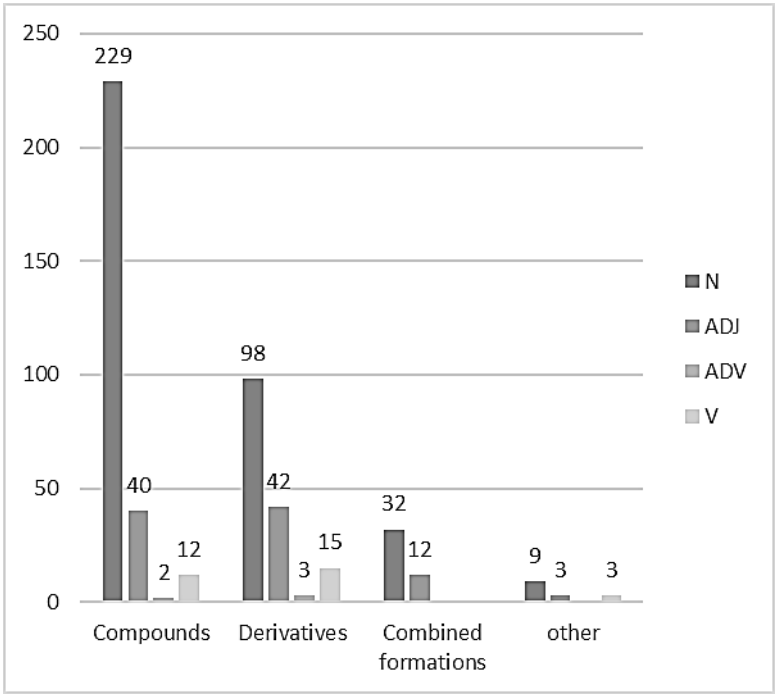

Figure 8: Distribution of word-classes and word-formation processes within the OED sample

Figure 8 shows that when lexical idioms are chosen primarily based on semantic anomaly, the distribution of word-classes and word-formation processes is significantly different from the general distribution of word-classes and word-formation processes (cf. Table 5 and Table 7 for the BNC sample). It can be seen that the proportion of nouns, in particular compound nouns, is very high. This is in accordance with expectations based on the fact that most accounts of idiomaticity in word-formation discuss indeed the category of compound nouns. However, the proportion of derivatives is definitely not negligible, which can be seen also in Table 24. On the other hand, some word-class categories are marginal, especially adverbs, and surprisingly also verbs are less frequent in the sample.

The data in the following analysis are structured according to the type (and possible subtype) of anomaly.

6.2.3. Semantic anomaly

As has been explained before, semantic anomaly is the primary criterion in this part of the analysis and therefore all 500 lexemes display some degree of semantic anomaly. Semantic anomaly is again, as in the BNC sample, seen in two different stages: polysemous lexemes

with both non-idiomatic and idiomatic senses, and lexemes with only idiomatic sense(s). The following tables (Table 25-27) illustrate the formal differences between the two groups:

	OED COMPLEX LEXEMES	
	NUMBER	%
TRANSPARENT AND IDIOMATIC MEANING	161	32.2
IDIOMATIC MEANING ONLY	339	67.8
TOTAL	500	100.0

Table 25: Proportion of stages of idiomaticity within the OED sample

	TRANSPARENT AND IDIOMATIC MEANING		IDIOMATIC MEANING ONLY	
	NUMBER	%	NUMBER	%
N	96	59.6	272	80.2
ADJ	51	31.7	46	13.6
V	11	6.8	19	5.6
ADV	3	1.9	2	0.6
Total	161	100.0	339	100.0

Table 26: Comparison of two stages of idiomaticity – word-classes

	TRANSPARENT AND IDIOMATIC MEANING		IDIOMATIC MEANING ONLY	
	NUMBER	%	NUMBER	%
COMPOUNDS	77	48.4	216	63.7
DERIVATIVES	68	42.8	83	24.4
COMB. FORMATIONS	14	8.8	25	7.4
OTHER	0	-	15	4.4
TOTAL	**161**	**100.0**	**339**	**100.0**

Table 27: Comparison of two stages of idiomaticity – word-formation processes

It can be seen that lexemes with only idiomatic meaning are more numerous within the sample. In addition, the distribution of word-classes and word-formation processes slightly differs in the two categories. The category of both idiomatic and transparent meaning contains a higher proportion of adjectives and the category of idiomatic-only lexemes includes a higher proportion of nouns. Derivatives are more common in the former category, whereas compounds in the latter category.

Apart from the differences described above, the lexemes were also studied in terms of their specific semantic subtype[9]. The subtypes found in the sample are described below and the range of possible semantic subtypes is the same for both categories discussed above.

6.2.3.1. Specialization of meaning

Assigned to this subtype are lexemes in which the lexical meaning is included in the word-formation meaning and the word-formation meaning is therefore wider than the lexical meaning. Examples of this phenomenon are listed in (31)

(31) *output, get-out, let-out, mobilization, activist, booklet, post-war*

The first three examples are examples of particle compounds with a low degree of idiomaticity (see also section 6.2.5.2. for further discussion of particle compounds). This kind of particle compounds (and their corresponding phrasal verbs) have a relatively transparent structure with no collocational anomaly between the two components. However, the verbs used are very vague and the lexical meaning is more specialized in comparison with the word-formation meaning. This is exemplified in Table 28 below:

PARTICLE COMPOUND	WORD-FORMATION MEANING	LEXICAL MEANING
output	"an action of putting something out"	"the act of production or manufacture"
get-out	"an action of getting out of somewhere"	"an escape, as from a difficult situation"
let-out	"an action of letting something out"	"a chance to escape"

Table 28: Comparison of word-formation and lexical meaning of selected phrasal verbs

[9] The examples are sometimes used with selected senses only to illustrate the category discussed although they may have other senses in addition to those mentioned in the text.

Note that *output* has also other meanings, the most common ones include "the amount produced", "the material produced" and "the information produced by a computer". These are not mentioned in Table 28 because they illustrate another type of semantic shift, i.e. metonymy (cf. 6.2.3.2.).

Mobilization "the act of preparing for war or other emergency by organizing (national resources, the armed services, etc.)" and *activist* "a person who works to bring about political or social changes by campaigning in public or working for an organization" are examples of semantic specialization in which the more specific meaning overrides the word-formation meaning (i.e. "the act of putting sth. into motion" for *mobilization* and "a person who is active" for *activist*). Both these examples have the word-formation meaning as one of their actual senses (although the word-formation meaning of *activist* is only listed in OED, not in CD), but the lexical meaning is more common. It must be also added that the specified meaning is present already in the base for *mobilization,* i.e. *mobilize*. Therefore, this formation would be considered regular if we only focused on the last step of word-formation and compared the lexical meaning to the word-formation meaning of *mobilize* + *-ation*.

Similarly *booklet* has the word-formation meaning "small book" and lexical meaning "small book containing information about something" (i.e. the purpose is specified). However, lexical meaning of this lexeme seems to be closer to the word-formation meaning as the informational character is only a typical feature and not a necessary criterion and the word-formation and lexical meaning are very close to each other. The degree of idiomaticity is lower in comparison with the previous examples.

Post-war (adj.) can be normally used in the word-formation meaning ("happening or existing after a war"), but it can be also used in the specified sense "after the Second World War, 1939-45". The specification is rather due to extralinguistic reality than internal factors and, therefore, this lexeme has the lowest degree of idiomaticity from the examples mentioned.

The main question associated with meaning specialization is to what extent it can be actually considered anomalous, as narrowing of meaning expressed by the form in comparison with its referent is a common phenomenon already at the point of coining new words and also during the process of lexicalization because both textual and situational context always narrow the

meaning to some extent. Bauer (1983: 57) discusses the lexeme *typewriter* explaining that the range of information not present in the word-formation meaning is broad, from the basic information about whether the referent is a person or machine, to the presence of the parts of the machine such as keyboard, shift-keys, etc. He points out that "it is not clear what the 'addition of semantic information' should be taken to include […] The most obvious answer might be to include […] only such information as is obligatory for the appropriate use of the word in referring, but the person/machine problem with *typewriter* suggests that even this is excessive." The issue is problematic also from the point of view of phraseology which focuses on the presence or absence of semantic anomaly. Is it really plausible to describe a common feature of lexicalization as an anomaly? In my opinion the criteria applied to single-word idioms must be adapted in this respect to the fact that the average degree of regularity on the level of words is considerably lower than on the level of word combinations. Čermák (2007b: 230) is aware of this problem mentioning depletive (i.e. vague) components which may cause opaqueness without being overtly anomalous (Čermák mentions *jack-pot* in this connection, but we could also add most particle compounds to this category, especially those which are to some degree transparent, such as *output* mentioned above). However, from another point of view, these lexemes can be seen as anomalous and thus included in the category of lexical idioms, as only those lexemes which are transparent (i.e. specific enough to identify their referent) can be seen as regular combinations with their word-formation meaning available for decomposition. Combinations including vague components, on the other hand, must be understood by their overall meaning only and must be seen as idiomatic units. This seems to be valid for lexemes with highly vague components, but there is certainly a large fuzzy area of less underspecified lexemes (where the specification by context may, or may not, be sufficient). We may therefore assume that it is useful to include meaning specialization into the possible discrepancies exhibited by lexical idioms, but at the same time meaning specialization should be considered as a borderline case of idiomaticity in which it depends largely on the degree to which the meaning is actually specified.

6.2.3.2. Metonymy

Metonymy is another common type of relation between the word-formation meaning and lexical meaning described in literature. It is based on an association between two entities

within the same domain. Some of the main types of these associations are presented in Nerlich (2006: 109): cause – effect, acts – major participants, part – whole, container – content, experience – convention, possessor – possession. There are 35 instances marked as metonymy in the OED sample. Here are some of them:

(32) *white-collar, face-to-face, superpower, chairperson, airfield*

(33) *footage, foreword, instrumentation*

(34) *outgrowth, word processor*

Compounds are exemplified in (32). Since compounds contain two lexical bases, there are two possible types of metonymy. The first type includes compounds which are metonymical as a whole, e.g. *white-collar* (adj.) "of, relating to, or designating nonmanual and usually salaried workers employed in professional and clerical occupations" is a metonymical transfer possessor – possession. *Face-to-face* is an instance of synecdoche (phrasal compounds are described below as a kind of formal anomaly). *Superpower* "an extremely powerful state" is a shift from attribute to possessor. The second type is characterized by a metonymical shift in one of the components: *Chairperson* is "a person who presides over a company's board of directors, a committee, a debate, an administrative department, etc.", i.e. *person* is used regularly, but *chair* is a metonymical shift of the type object – function. Similarly, *airfield* "a landing and taking-off area for aircraft" has one regular component, *field*, and one metonymic component, *air*, shifting between locality and occupant.

Examples in (33) illustrate metonymical derivatives. *Footage* "the sequences of filmed material" is based of metonymical shift from the extent of film tape (measured in feet) to the length of the filmed sequence or its position on the tape (the reference to digitally recorded material with no tape involved can be seen as meaning generalization due to changes in extralinguistic reality). *Instrumentation* is used either regularly as an abstract noun "the use of instruments" or metonymically as a collective noun "the instruments specified in a musical score or arrangement". *Foreword* "introduction in a book" is an instance of synecdoche, referring to the whole text by its part, *word*. This last example is similar to *chairperson* and *airfield* in having only one part metonymical: *fore* is used with its regular meaning.

The last two examples in (34) illustrate metonymical combined formations. *Outgrowth* can be regular in "the act of growing out", or metonymical in "a thing growing out of a main body" (the third, metaphorical, meaning will be dealt with below). The second example, *word processor* has only the first component metonymical, *processor* is used regularly (with slightly specialized meaning).

From the theoretical point of view, it seems that the latter type of metonymy, the combination of regular and metonymical components, is more idiomatic than the former type, if we understand lexical idioms as anomalous combinations of sublexemic components. In these cases, the anomaly is already present at the moment of coining, whereas in the former type the metonymical meaning may develop later from a perfectly regular meaning (although there are instances such as *face-to-face* where the metonymical meaning will have arisen also already during the moment of coining).

Nevertheless, it is rather problematic to analyse metonymical relations and their origin, because metonymy is a common meaning shift used on several levels of language production (rhetorical figure, lexicalized simple words, lexicalized complex words) and cognitive linguistics sometimes describes it as a conceptual phenomenon (similar to conceptual metaphor, cf. Kövecses and Radden, 1998). Since metonymy is such a ubiquitous phenomenon in semantics, it is also problematic to decide whether examples such as *chairperson* or *schoolboard* should be seen as anomalous combinations of a shifted and regular component, or whether we should take into account that both *chair* and *board* are already used in the metonymical sense as simple lexemes and consider therefore the compounds to be non-idiomatic.

In summary, since metonymy displays some systematicity and it is rather frequent in language use, it seems reasonable to describe it as a semantic discrepancy with a low degree of idiomaticity, especially if it is not combined with any other anomaly.

6.2.3.3. Metaphor

Metaphor is usually defined as semantic shift based on similarity or analogy which is not explicitly expressed (which differentiates metaphor from simile). Metaphor is approached differently in various theories and it is, on the one hand, often associated with idioms, cf.

Howarth's (1998) figurative idioms which are based on metaphors and Burger's (1998: 83) metaphorical idioms, i.e. lexicalized metaphorical phrases, and, on the other hand, it is treated by cognitive linguists as a systematic phenomenon, especially in the conceptual metaphor theory (cf. Benczes, 2006, who argues for considering creative metaphorical compounds as transparent because of the common knowledge of the underlying conceptual metaphor).

Lexemes based on metaphor constitute a significant part of the OED sample. Some of them are exemplified below:

(35) *multidimensional, encapsulate, telling*

(36) *break-off, viewpoint, blueprint*

(37) *feedback, work load, side effect, pinpoint, dead end*

(38) *outgrowth, overriding, broadcaster, pacemaker, fundraising*

Lexemes in (35) are metaphorical derivatives. The first three examples are used in both literal and metaphoric meaning. In the first two instances, *multidimensional* and *encapsulate*, the metaphor is fairly transparent: the non-physical interpretation of *dimension* is a common process of mapping abstract ideas on concrete objects and presenting ideas as something which can be *encapsulated* is similar in this respect, drawing on the conceptual metaphor IDEAS ARE FOOD (cf. Lakoff and Johnson, 1980). *Telling* (adj.) is different in having only the metaphorical meaning "revealing" or "having a marked effect or impact". This makes *telling* the most anomalous. The semantic shift differentiates the converted adjective from the participle form, which is regular.

Compounds in (36) are used in both literal and metaphorical meaning. *Break-off* "the act of breaking off" or "an abrupt discontinuance, especially of relations" is again quite transparent as an instance of concrete – abstract shift, but the additional specialization of its reference is less transparent. The abstract meaning of *viewpoint* (which is more common than the concrete, literal, meaning) is also quite transparent, being based on the conceptual metaphor UNDERSTANDING IS SEEING. *Blueprint* seems to be more idiomatic because the regular meaning "a photographic print of plans, technical drawings, etc., consisting of white lines on a blue background" and the idiomatic meaning "an original plan or prototype that influences

subsequent design or practice" are linked by the idea of a template which is reproduced repeatedly. Unlike in *viewpoint,* the relevant facet of the literal meaning is not directly retrievable from the word-formation meaning of *blueprint* (i.e. knowing the meaning of *blue* and *print* does not help in decoding the metaphorical meaning).

Examples in (37) are used only in the metaphorical sense: *feedback* is used either in one of the specialized technical senses involving "the return of a part of the output to the input" or in the general sense "information in response to an inquiry, experiment, etc." Neither of these is directly based on the primary sense of *feed* associated with *food*. The second component, *back* is used regularly (general properties of particle compounds are dealt with in 6.2.5.2.). *Workload* and *side effect* are examples of nominal compounds with only one component metaphorical: *load* in the former and *side* in the latter lexeme. *Load* is a transparent shift from concrete to abstract, whereas *side* in the meaning "unwanted, secondary", referring to importance and function of something in terms of spatial position, may not be directly obvious. *Pinpoint* "to locate or identify exactly" and *dead end* "cul-de-sac" or "a situation in which further progress is impossible" are instances of high idiomaticity, which is caused by an uncommon metaphorical shift (accompanied by a formal or collocational anomaly, which will be described below in 6.2.4.).

Examples provided in (38) are instances of combined formations with metaphorical meaning shift. We have found in the BNC sample that combined formations tend to be more often transparent due to their complex onomasiological structures. Nevertheless, they can be idiomatic if the whole formation or a part of it is used metaphorically. *Outgrowth* has been discussed above as an instance of metonymy. It can be also used metaphorically in the sense "a result; consequence; development" in accordance with the conceptual metaphor IDEAS ARE PLANTS. *Overriding* "taking precedence" is used in this sense (along with several related senses) already as the verb *override.* Nevertheless, neither verb nor adjective are used in the word-formation sense based on the combination of *over* and *ride*. *Broadcaster* is based upon a verb, *broadcast*, which is metaphorical itself, and the agentive suffix added in the last step is regular. *Pacemaker* is used regularly (with some degree of meaning specialization) in "a person, horse, vehicle, etc, used in a race or speed trial to set the pace", but metaphorically in "a person, an organization, etc, regarded as being the leader in a particular field of activity".

The metaphorical use is based on a fairly transparent conceptual metaphor BUSINESS IS JOURNEY. In addition, it has a specialized medical meaning (metaphorical as well). *Pacemaker* differs from *broadcaster* in being a synthetic compound in which all three parts are usually understood to be combined simultaneously, and therefore, the idiomatic meaning (based on the conceptual metaphor) is first associated with the compound. *Fundraising* has a semantically anomalous head *raising*, which is only used in the sense of "collecting" when used with money.

It can be seen from the examples above that metaphor is further on the scale of idiomaticity than metonymy because metaphors are domain shifting and the word-formation and lexical meanings are therefore at least superficially more distant from each other. Although the cognitive approach argues for viewing metaphors as systematic, based upon general concepts which reflect our common understanding of the world, it seems reasonable to accept that phraseology, focusing on combinatorial anomalies, views the use of components from another domain as semantically anomalous because that is what they are in comparison to the regular, literal, use of the components. Nevertheless, there is a cline with less idiomatic instances of common conceptual metaphors and more idiomatic instances of less obvious metaphors (often accompanied by another anomaly, such as further meaning specialization as in *break-off* or semantic incompatibility in *dead end*).

6.2.3.4. Exocentric formations

As has been pointed out in 5.3., exocentricity is a major semantic anomaly which can be described as an instance of figurative language use. Fifty lexemes in the sample are marked as exocentric, all of which are nouns.

The main types of exocentric compounds are exemplified below:

(39) *know-nothing, paperback, close-up, high-rise, desktop, freelance*

(40) *print-out, dug-out, leftover, spin-off, hangover*

The first noun from (39), *know-nothing*, is a compound traditionally designated as the Romance type having the syntactic head to the left. Although the lexeme is an instance of exocentric formation, and the components do not explicitly mark the referent, i.e. "an ignorant person", it is in fact quite transparent because of its semantics (*knowing* is typical of persons,

not objects). Nevertheless, even though the formation is not very opaque, it is still reasonable to consider it a lexical idiom due to its exocentricity which must be always considered as an unexpected semantic anomaly. *Paperback* is a possessive compound (the bahuvrihi type) referring to "a book with a thin cardboard or paper cover". In addition to the exocentricity, *back* is anomalous in describing the cover of a book. *Close-up*, "a photograph or film or television shot taken at close range", is formed by conversion from an adverbial multi-word idiom *close up*. In this aspect it is similar to examples in (40) discussed below. *High-rise* and *desktop* are instances of word-ellipsis based on multi-word expressions *a high-rise building* and *a desktop computer*. The last example in (39), *freelance* "a self-employed person", is surely the most idiomatic as there is no transparent relation between any of the components and the referent (which is caused by diachronic semantic shifts).

Examples in (40) are all instances of the same formal type, particle compounds. These formations are all exocentric because they consist of a phrasal verb (verb + particle) converted into a noun. Therefore, the formal head is verbal and not nominal. There are less idiomatic instances within the category, such as *printout* "a piece of paper on which information from a computer or similar device has been printed", which refers to a prototypical instance of "something printed out" or more idiomatic instances such as *dugout* "a canoe made by hollowing out a log" which refers to a less obvious product of the action described by the correspondent phrasal verb. *Leftover* "an unused portion or remnant, as of material or of cooked food" is formed without much change from the participial collocation *left over*, although the non-participial form *leave over* is rare. *Spin-off* is more idiomatic than the preceding instances because the verbal meaning "to produce as an outgrowth or secondary benefit, development, etc." is only partly corresponding to the senses of the noun "any product or development derived incidentally from the application of existing knowledge or from an enterprise" or "a book, film, or television series derived from a similar successful book, film, or television series". *Hangover* is similar, the nominal meaning "the delayed aftereffects of drinking too much alcohol" includes a lot of meaning specialization in comparison with the verbal meaning "to be left from a previous time or state". The second nominal meaning "a person or thing left over from or influenced by a past age" corresponds to the verbal meaning.

Exocentric compounds are not the only exocentric category found in the sample. There are also instances of derivatives which may be described in terms of exocentricity. Derivatives are not traditionally described as exocentric in literature although Bauer (2016: 467-469) discusses the possibility of their inclusion into the category. He provides the example *revolver*, which does not describe "an instrument which revolves" but "an instrument which has a part which revolves". However, Bauer also points out that once we admit the existence of the category, we may include a large number of common lexemes based typically on metonymy such as *carriage* "vehicle", *diner* "restaurant", etc. Bauer (p. 469) concludes that "[w]hat we see in all these cases is an instance of derivational word-formation which is not obviously compositional for the very same reasons that so-called exocentric compounds are not compositional. The reasons are connected to a figurative interpretation of some kind. To the extent that there are exocentric compounds, it would seem reasonable to suppose that there are also exocentric derivatives, but these do not feature in the literature."

The OED sample contains 10 derivatives which are marked as exocentric, but systematic cases of metonymy of the type action – instrument (*carriage*), action – product (*building*), etc. were not included, although they are probably only slightly more systematic instances of the category. Here are several examples of the category:

(41) *detective, documentary, floppy, microwave, surround*

Detective "a police officer who investigates crimes" has an adjectival suffix *-ive* and the regular meaning would be adjectival "serving detection". The nominal meaning is an ellipsis from *detective policeman.* The three following examples are also instances of ellipsis: *documentary* for *documentary film*, *floppy* for *floppy disk*, *microwave* for *microwave oven.* The latter instance is different in not having a word-class distinctive suffix, but a prefix which does not serve as a category marker. By this it is closer to exocentric compounds in having a lexical base as the syntactic head. The last example, *surround* (n.) "a border, esp. the area of uncovered floor between the walls of a room and the carpet or around an opening or panel" is an instance of conversion describing "something that surrounds". Nevertheless, conversion may be seen as a fairly systematic instance of exocentricity and because of this fact it is less idiomatic than instances of univerbation by ellipsis.

A question may arise whether other word-classes also include lexemes which may be classified as exocentric. Looking for exocentricity in other word-classes is rather problematic because the non-hyponymic relation of the referent to the syntactic head (which is the common definition of exocentric compounds) is more difficult to assess. Adjectives, for instance, contain a number of compounds such as *high-level*, *post-war* or *pre-term* which might be seen as exocentric as they have a nominal head. However, these adjectives are usually used in the attributive position, where noun modifiers regularly occur, and there is nothing anomalous in them in this respect. It would probably sound more plausible to talk of exocentricity in connection to adjectives with formal features of other word-classes, such as deverbal *go-ahead* or deadverbial *far-out*. As far as verbs are concerned, the most typical examples appear to be those with a suffix distinctive of a different word-class (which is also mentioned by Bauer, 2016). The sample includes verbs *package, layer* and *buffer*; or instances of verbs converted from nouns with clearly nominal head such as *highlight*, *streamline*, *pinpoint*.

Nevertheless, in may be said in conclusion that although exocentricity may lead to higher opaqueness in some instances and lower opaqueness in others, it is always anomalous at least formally because exocentric formations do not signal the nature of the referent by their formal structure.

6.2.3.5. Minor semantic subtypes

The previous sections described four most frequent discrepancies between word-formation and lexical meaning. Apart from them, the sample provided sporadic instances of other types described below:

(42) *suitcase, poster, used-up, handbag*

(43) *exceptional, resourceful, standout*

(44) *collaborator, mechanistic, opportunist*

Examples in (42) are instances of generalization of meaning, i.e. the lexical meaning is wider and includes the word-formation meaning: *suitcase* is not only for suits but also for other clothes and items, *poster* is not distributed only by post, *used-up* means not only “consumed completely” but also “exhausted, worn out”, *Handbag* is a case of both generalisation and

specialization relating to two different aspects of meaning, being defined as “a woman's small bag carried to contain personal articles”. The lexical meaning is more general in that it is not only held by hand, but also worn over the shoulder, and more specialized in being restricted to women’s bags.

Lexemes in (43) are instances of amelioration of meaning: *exceptional* means not only “forming an exception” but also “having much more than average intelligence, ability, or skill”. *Resourceful* is not only “full of resource” but also “ingenious, capable, and full of initiative” and *standout* does not mean “a person standing out” but “a person or thing conspicuously superior or notable in performance, quality, etc.” (the meaning is metaphorical in addition).

The last group in (44) are instances of deterioration of meaning: *collaborator* is not only “a person working together with someone else” but also “a person who helps an enemy who is occupying their country during a war”, *mechanistic* means not only “of or relating to mechanics”, but it is also an evaluative expression criticizing someone for describing a natural or social process as if it were a machine. *Opportunist* is not related regularly to *opportunity*, but it is again an evaluative term criticizing “a person who adapts his or her actions, responses, etc, to take advantage of opportunities”.

There is also an instance of commonization in the sample (*narcissistic*), instances of euphemisms (*developing* “poor, underdeveloped”), instances of determinization (e.g. *exponentially* “very rapid”, *fluorescent* “glowing and vivid”) and hyperbole (*exhausted* “very tired”).

These minor subtypes are often linked to evaluative function and the discrepancy between the word-formation and lexical meaning often involves the connotative component of meaning. Changes in connotation are often rather unstable and therefore the sense may change relatively rapidly. This may be illustrated by the lexeme provided above, *exceptional*, which has an additional meaning in American English “needing special attention or presenting a special problem, as in education, because mentally gifted or, esp., because mentally, physically, or emotionally handicapped”. We can therefore see that the transition between meaning amelioration and deterioration is sometimes quite easy.

6.2.3.6. Combination of more subtypes of semantic anomaly

It has been indicated in connection with some words mentioned in the previous sections that the above subtypes may also combine. In such cases, the degree of idiomaticity probably rises as the relation of word-formation and lexical meaning is more complex.

Some of the combinations are exemplified below:

(45) *airport, outgrowth*

(46) *sketchy, big brother, one-dimensional*

(47) *turnaround, standout*

Examples in (45) combine metonymy and metaphor: *airport* is composed of metonymical *air* and metaphoric *port*, *outgrowth* describes a result of growing (metonymy) in the metaphorical sense "a consequence".

Lexemes in (46) exemplify a combination of metaphor and deterioration of meaning: *sketchy* "lacking completeness; rough; inadequate" is at the same time metaphoric and has negative connotation which is not part of the word-formation meaning. *One-dimensional* may refer to something "having a single focus; narrow and superficial" with negative connotation and b*ig brother* "a person, organization, etc, that exercises total dictatorial control" is metaphoric in likening something to the concept known from Orwell, at the same time expressing negative connotation.

Lexemes in (47) are a combination of metaphor and amelioration: *turnaround* in one of its meanings, "a sudden improvement, especially in the success of a business or a country's economy" is used figuratively with a positive aspect of the lexical meaning not given by the word-formation one, and *standout*, described above, is in addition to metaphoric and ameliorative aspect also exocentric.

6.2.3.7. Semantic anomaly of unspecified type

Apart from the subtypes of semantic anomalies described above, there are also lexemes in the sample which are non-compositional, but the semantic discrepancy cannot be described as an instance of the above-mentioned categories. In fact, a large proportion of the sample belongs to this unspecified category. It is probably useful to take a closer look at these cases. I make

no attempt to sort the lexemes from this group into categories because there are just too many idiosyncrasies, but we can still find lexemes which are in some aspects similar to each other. These are discussed below.

One group of these lexemes consists of particle compounds based on more idiomatic phrasal verbs. This group differs from particle compounds in example (31) in that it is difficult to see the relation between their lexical and word-formation meaning: lexical meaning is not just more specified word-formation meaning, it is widely divergent. Several such nominal compounds are exemplified in (48):

(48) *make-up, back-up, workout*

Make-up meaning "face cosmetics" is opaque because the word-formation meaning (*make* + *up*) has little to do with the lexical meaning. The same applies to *back-up* "additional support or resources to help accomplish a task" (and other senses of the noun) and *work-out* "a period of physical exercise or training". Even though we attempt to classify particle compounds according to their semantics as fairly transparent exocentric compounds (such as *printout* discussed above), fairly transparent instances of meaning specialization (such as *let-out* discussed above) and opaque formations (such as *make-up*), it must still be emphasized that the classification is very subjective especially because of the vagueness of the components, which is typical of phrasal verbs. Particle compounds and their relation to phraseology will be discussed later in 6.2.5.2.

It is also useful to look at some specific subgroups inside this large group. If we look at lexemes with only semantic anomaly (i.e. without additional formal or collocational anomaly), we may find more about the properties of these idiomatic lexemes.

Let us look at derivatives first:

(49) *reactor, organizer*

(50) *connectivity, operational*

(51) *subtitle, phenomenal, hippie*

Examples in (49) illustrate anomalous semantic relations of the base to the suffix: in *reactor* "a vessel, esp. one in industrial use, in which a chemical reaction takes place" and *organizer*

"a container with a number of compartments for storage", the regular (and most productive) relation of base and *-er*/*-or* is agentive or instrumental, i.e. the base is verbal, and the meaning is "the person or thing that does the activity expressed". Formations of this type (with a verbal base) are formally regular, but semantically anomalous, whereas formations with nominal base are classified here as both semantically and formally anomalous.

The two lexemes in (50) illustrate anomalous meaning of affixes: in *connectivity* "the state of being connected to the internet", it is the adjectival *-ive* which is anomalous because its basic meaning "tending to - " (Bauer and Huddleston, 2002: 1711) is not present in the lexical meaning of *connectivity* in the sense provided above. In contrast, *operational* "in working order and ready for use" is anomalous because *-al* expresses here the meaning which is normally covered by *-able* or *-ive* (i.e. tendency, capability).

Examples in (51) have semantically anomalous bases: *subtitle* "a written translation superimposed on a film that has foreign dialogue" or "explanatory text on a silent film" has an anomalous base, *title,* and *phenomenal* "extraordinary; outstanding; remarkable" does not relate directly to *phenomenon* but has undergone a semantic shift. *Hippie* is an instance of opaque formation due to diachronic reasons: it is usually (e.g. in the OED) associated with *hip* "well informed, in the know". However, the common sense of the word has shifted to "modern, fashionable", which makes the base even more opaque. In addition, the formation is very vague, which increases the overall opacity. Most of these examples (with the exception of *hippie*) are also used in the regular, non-idiomatic, meaning, although it is usually not the more frequent use. This implies that they have gone through a subsequent semantic change, acquiring a new sense besides the regular one.

We may now turn to compounds and inspect them in a similar manner. They are arranged in groups corresponding to those distinguished in derivatives:

(52) *hallway, timeline, bureaucracy*

(53) *check-list, runway, grandparent*

(54) *screenplay, windshield, black box*

Lexemes in (52) exemplify compounds with a semantically anomalous head, but regular modifier: *hallway* may refer not only to "corridor", but also to "hall" in British English. In such a case, the head *way* is anomalous as the noun refers not to a *way*, but rather to an open space. *Timeline* denoting "a time frame during which something is scheduled to happen" is related to *time*, but it is not a *line* in its regular sense (the lexeme is non-idiomatic in the sense "a graphic representation showing the passage of time as a line"). *Bureaucracy* standing for "a system of administration based upon organization into bureaus, division of labour, a hierarchy of authority, etc: designed to dispose of a large body of work in a routine manner" and its related meanings, takes its first part from *bureaux*, but it is not "a type of government" as the combining form head *-cracy* specifies.

Examples in (53) illustrate compounds with an anomalous modifier and transparent head: *check-list* is not restricted to *checking*, defined as "a list of items, facts, names, etc, to be checked or referred to for comparison, identification, or verification", *runway* "a hard level roadway or other surface from which aircraft take off and on which they land" is a kind of way, but not for running, and *grandparent* is a type of parent but with an opaque modifier (ultimately deriving from Latin *grandis* 'full-grown').

Lexemes in (54) represent idioms with the highest degree of idiomaticity because they are anomalous in both parts. *Screenplay* "the script for a film, including instructions for sets and camera work" has neither direct relation to *screen*, nor to *play.* It was coined as a term (for discussion of terms, see 6.2.6.) based probably on some metonymical shifts in both parts. *Windshield* "the sheet of flat or curved glass that forms a window of a motor vehicle, esp. the front window" is idiomatic due to changes in extralinguistic reality: the original front window was in roofless cars where the front glass really served mainly as a wind shield. Nevertheless, the protective function from wind is nowadays less prominent since the design of cars has changed. *Black box,* in both its common senses "an electronic device in an aircraft which records information about its flights" or "a self-contained unit in an electronic or computer system whose circuitry need not be known to understand its function", does not have to be *black* and the polysemous head *box* is too vague.

As far as combined formations are concerned, there are only few examples of semantically anomalous lexemes which are not specified in terms of subtype as most instances of idiomatic combined formations are based on metaphor or metonymy. They are provided below:

(55) *loudspeaker, bestseller, smallholder, air conditioner*

Loudspeaker "reproducer" is idiomatic by combining three parts, none of which is used in its prototypical meaning: *speak* is prototypically used with agentive nouns, but not in this case, suffix *-er* refers to an instrument (not agent) although it combines with *speak,* and *loud* is anomalous as well (it has rather the meaning of *aloud*). *Bestseller* "a book, record, CD, or other product that has sold in great numbers, esp. over a short period" again combines several anomalous aspects: it is partly based on meaning specialization (it is normally used to refer to only some specific products) and it is partly idiomatic by using *sell*, which is typically combined with an agent of the action*,* with the suffix *-er* employed in neither agentive nor instrumental function (a *bestseller* is the patient of the action of selling). *Smallholder* "a person who owns or works a smallholding" may be seen as opaque because the suffixation changes the superficial structure: the correct word-formation analysis is [[small - hold] - er], but since **smallhold* is a formally irregular bound base (back-formation from *smallholding*), the structure can be reinterpreted as an opaque structure [small - [hold - er]] which is a common ADJ + N structure in compounds. *Air conditioner* is a similar case: *air-conditioning* served as a base for back-formed *air-condition*, which resulted later in later *air conditioner.* However, *air conditioner* can be reinterpreted to become opaque [air + [condition + er]]. We can conclude that the last two instances may be seen as cases of formal anomaly (due to back-formation) which leads to ambiguity in semantic decomposition.

6.2.3.8. Semantic subtypes – quantitative data

The following table summarizes the distribution of semantic subtypes in the sample. A lexeme was included into one of the categories only if the discrepancy between the word-formation and lexical meaning is based primarily (or solely) on one (or more) of the semantic subtypes studied. If the lexeme contains an additional idiosyncratic semantic shift, it is included in the first category, marked *no specific subtype*. Some lexemes are included in two or more categories if the shift includes more of the processes studied, and therefore the sum of

all subtypes do not correspond to the number of lexemes in the sample. It must be taken into account that the numbers are only rough as the delimitation of most of the categories is somewhat arbitrary and there is a large fuzzy area of lexemes with some aspects of these categories combined with an idiosyncratic semantic shift.

SEMANTIC SUBTYPES	OED SAMPLE OF IDIOMATIC COMPLEX LEXEMES	
	NUMBER	%
no specific subtype	165	33.0
specialization	123	24.6
metaphor	105	21.0
exocentricity	50	10.0
metonymy	35	7.0
generalization	10	2.0
deterioration	8	1.6
amelioration	6	1.2

Table 29: Distribution of semantic subtypes within the OED sample

It can be seen in Table 29 that apart from the large group of lexemes with unclear semantic relations between word-formation and lexical meaning, it is meaning-specialization which is most numerous in the sample. However, it has been mentioned in 6.2.3.1. that this category is marginal in the field of lexical idioms. Therefore, it seems that it is indeed the commonly mentioned metaphor and exocentricity, which are the most typical semantic subtypes of lexical idioms, whereas metonymy, generalization, deterioration and amelioration are less typical.

Figures 9 to 12 below provide data about the most common subtypes combined with data about the respective word-formation process:

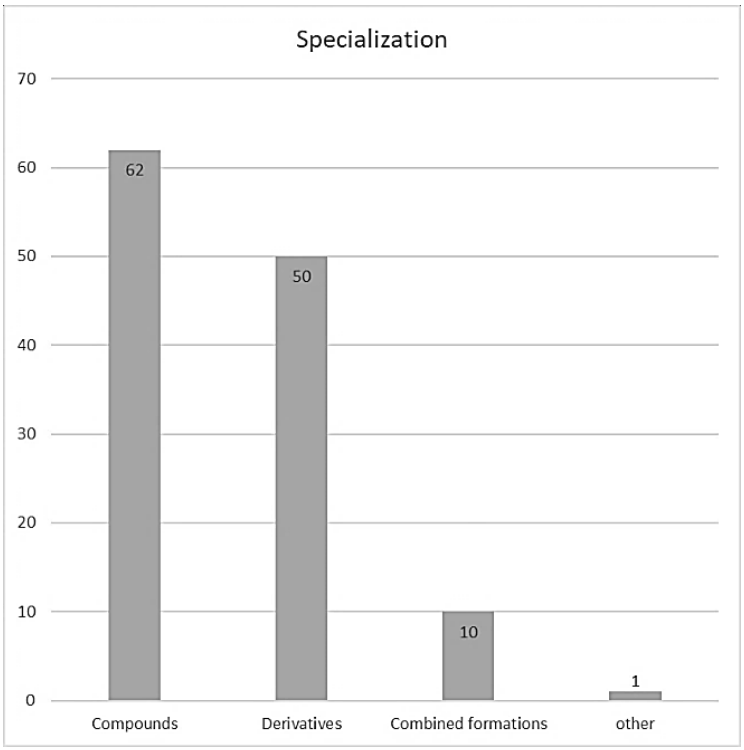

Figure 9: Distribution of word-formation processes in the subtype *specialization*

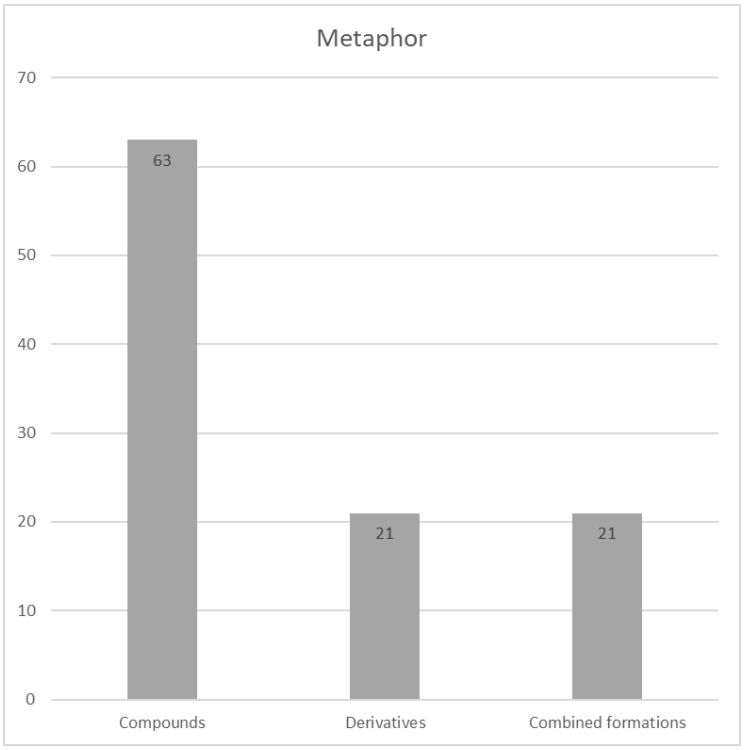

Figure 10: Distribution of word-formation processes in the subtype *metaphor*

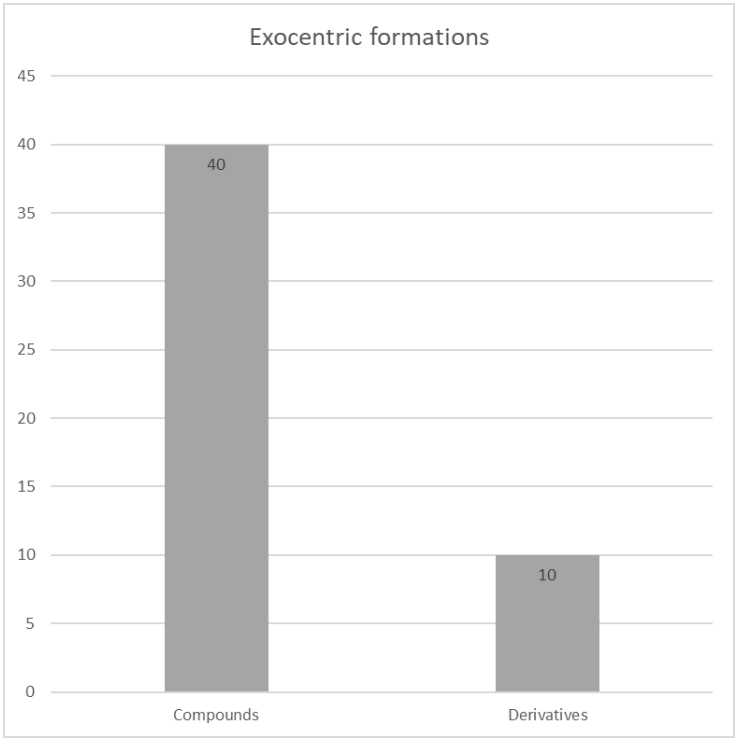

Figure 11: Distribution of word-formation processes in the subtype *exocentricity*

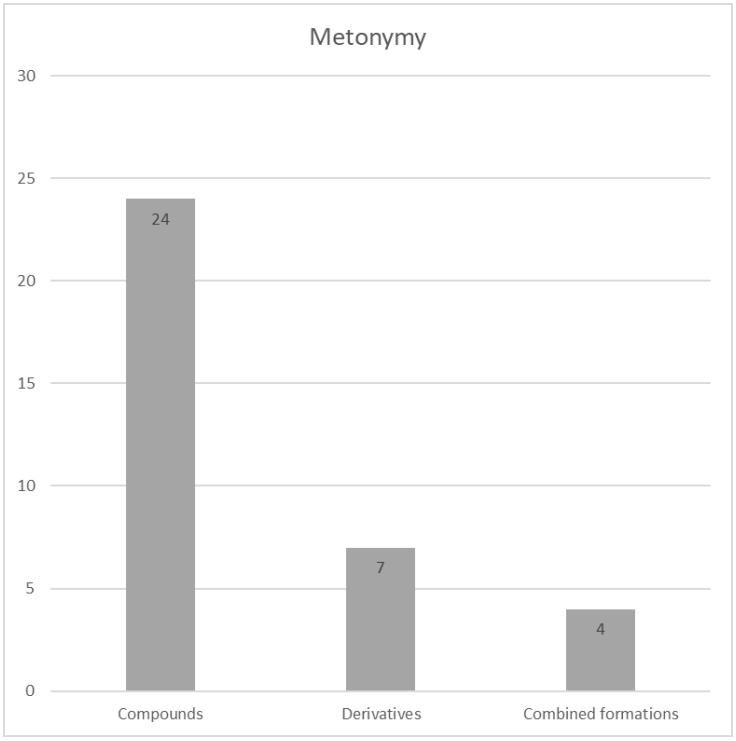

Figure 12: Distribution of word-formation processes in the subtype *metonymy*

6.2.4. Formal and collocational anomaly

This section describes lexemes in the sample which are, in addition to semantic non-compositionality, also anomalous in their form (formal anomaly) or in the way their components are combined (collocational anomaly). These two categories are dealt with in one section because they to a large extent overlap. Since the main type of formal anomaly in Czech lexical idioms discussed by Klötzerová (1997, 1998), i.e. non-productive word-formation processes, is disregarded in the present analysis of English lexical idioms (cf. 6.1.3.2.), the range of formal anomalies is rather restricted. One of the main types of formal anomaly is the anomalous combination of components, which represents a borderline class between formal and collocational anomaly. In the present analysis, it is the semantic anomaly which is considered primary, accompanied sometimes by formal and/or collocational anomaly. However, it seems that in some cases it is actually the formal or collocational anomaly which triggers also semantic anomaly. This is especially the case of rare components which are opaque due to their low collocability. In addition, there are infrequent lexemes in which components are formally anomalous, but not to such an extent that they should be considered semantically opaque. In such cases, the formal idiosyncrasy could be considered the only anomaly present in the lexical idiom. This is, for example, the case of *spokesperson* discussed below in (57) in which the formally irregular *spokes* might still be seen as transparent. These instances were nevertheless included in the sample as well because these formally anomalous components display low collocability and it would be questionable where to draw the line between the transparent and the non-transparent within this category. They are treated as peripheral instances of lexical idioms.

The following discussion of concrete types and examples starts with formal anomalies, proceeds to the borderline cases and ends with semantic incompatibility and redundancy in word-formation.

6.2.4.1. Formal anomaly

The first type of formal anomaly is described here as one involving an anomalous component form. The class includes lexemes which are not formed regularly from the common forms of

corresponding morphemes (or complex regularly formed bases), but the form of the components are somehow idiosyncratic.

The class is exemplified by these two groups:

(56) *contraceptive, cultivar, surfactant, catalyse*

(57) *T-shirt, spokesperson, handicapped*

Lexemes in (56) exemplify splinter formations. *Contraceptive* is a clipping derivative formed by the prefix *contra-*, splinter *-cept-* (from *concept*) and sufix *-ive*. *Cultivar* is a clipping compound formed from *cultivated variety. Surfactant* is a combined formation formed from *surface-active* + suffix *-ant* involving clipping as well. *Catalyse* is an instance of blending, formed from *catalysis* and *analyse*. Blend formations are in some aspects peripheral in the category of lexical idioms because they are typically semantically quite transparent, especially those formed as ad hoc formations since the intended effect depends on their being understandable, i.e. transparent.

Examples in (57) present another group of formally anomalous components, *T* in *T-shirt* is used iconically for its shape, and does not have any conceptual meaning. *Spokes* in *spokesperson* (by analogy with *spokesman, spokeswoman*), irregularly formed probably from *spokes* (possibly singular of *spoke*, past participle of *speak* used as a noun) + *man/woman/person*, has been discussed above. *Handicap* as the base of *handicapped* is a form amalgamated from the phrase *hand in cap*, which is however extremely opaque and it would be definitely possible to regard the base as a simple morpheme from the synchronic point of view. Nevertheless, since both *hand* and *cap* are recognizable in the formation, the lexeme has been included in the sample.

The second type of formal anomaly is the use of an anomalous suffix. Lexemes of this type have an affix which is typical of one word-class (and therefore signalizes the word-class), but the lexeme is actually of a different word-class. However, it is necessary to realize that English is different from inflectional languages such as Czech in this respect because formal signs of word-class are less important as they do not interfere with inflection. In addition, the proportion of simple lexemes is higher in English and frequent conversion, especially between nouns and verbs, causes that the perception of these signs as anomalous features is lower.

Because of these facts, only typical suffixes were included, although the category could be expanded to other instances of word-class signals, for example ADJ+N compounds could be seen as typically nominal, and therefore anomalous when representing other word-classes (e.g. adjectives *present-day, long-range, high-level* or the verb *highlight*). However, due to frequent conversion, the transition of one form between nouns and adjectives or nouns and verbs is quite common, and it is debatable whether these lexemes should be considered formally anomalous when they combine two regular word-formation processes (compounding and conversion). Affixes more or less typical of two word-classes, such as *-ing*, were not included. Several instances of anomalous affixes are exemplified in (58):

(58) nouns: *documentary, wireless, deductible;* verbs*: package, layer*; adjective: *high-pressure*

In all instances mentioned in (58), the anomaly is caused by conversion. The nouns with adjectival suffixes can be probably considered most idiomatic because they are all exocentric in addition and the meaning is quite opaque. Verbal instances of this class are less idiomatic because although they might be seen as exocentric as well (the action is not part of their onomasiological structure), the action meaning is quite transparent (*package* "to make package(s)", *layer* "to make layer(s)") and converted adjectives are least idiomatic because the process of using phrases in the attributive position as syntactic adjectives (*a high-pressure pump, a long-distance race)* is fully productive (they are included in the sample only if they show some additional semantic discrepancy).

One of the formal subtypes is represented by phrasal compounds, i.e. single-word units created not by word-formation processes in the narrow sense but by the freezing of a word combination and using it on a lower level of language structure, i.e. the process of univerbation. Some phrasal compounds from the sample are exemplified in (59):

(59) *day-to-day, no-good, must-see*

The last formal type, represented only by three lexemes in the sample, is syntactic/morphological anomaly:

(60) *value-added*, *overdue, overseas*

Value-added (N) is anomalous by the postposition of the adjective, whereas *overdue* is anomalous by its syntactic use: *due* is used in the relevant sense only predicatively, but *overdue* can be used attributively as well (where *due* alone has a different meaning). *Overseas* has a fossilized, archaic, genitive form *seas*, which makes it idiomatic in the present-day English.

6.2.4.2. Formally anomalous combination of components

Formally anomalous combinations are combinations which are not made according to the common combinatorial rules. Some instances of this type from the sample are exemplified as follows:

(61) *movie, steamer, capacitor, insider, fledgeling*

(62) *knowledgeable, consumerism, tailored, retiree*

(63) *ceasefire*

Lexemes in (61) are anomalous by an uncommon combination of a base and affix. *Movie* combines a verb with *-ie* although most formations with this familiarity marker are based on nouns (e.g. *doggie, hanky, girlie*) or adjectives (e.g. *baddie, indie*). This is due to the ellipsis of the head noun in the expression *moving picture,* with the premodifier assuming the function of a noun. *Steamer* is a combination of the nominal base and the suffix *-er*, which is not as common as the combination of this suffix with verbal bases. Although nominal bases do occur with *-er,* they can be considered anomalous in when compared to the far more productive agentive or instrumental deverbal formations. *Capacitor* is again a nominal base with *-er,* but this time the base already has an affix, which is even more anomalous. *Insider* combines an adverb with *-er* and *fledgeling* combines a verb base with the deminutive *-ling*, which is less common than the combination of this suffix with nominal or adjectival bases (although not completely uncommon, cf. *starveling, hatchling*).

Lexemes in (62) combine two affixes in an uncommon way: *knowledgeable* combines *-ledge* and *-able* in one lexeme, which only occurs according to the OED in this word and in *acknowledgeable.* In *consumerism* the native suffix *-er* is followed by the non-native *-ism*, which is a possible formation (cf. Bauer, Lieber & Plag, 2013: 594), but generally the possibility of a native suffix preceding the non-native suffix is uncommon in English.

Tailored is a formally marked nominal base followed by a verbal prefix (which can be explained by the fact that the suffixation was preceded by conversion). *Retiree* is a peripheral case which might be seen as either regular or idiomatic: it is formed regularly from the verb *to retire*, which would imply that *retiree* is the patient of the action (by analogy with *employee*), but *retiree* is normally understood rather as the agent (CD defines *retiree* as "a person who has retired from work" and OALD as "a person who has stopped working because of their age").

Ceasefire in (63) exemplifies the Romance type of compounds (cf. also *pickpocket, scarecrow,* etc.). Although these formations do occur, they are not very common and they are almost unproductive. Therefore, they present a word-formation anomaly in the English system of compounds.

A subtype of formally anomalous combinations is that in which one of the components is monocollocable or has extremely low collocability (i.e. occurs in just one or very few lexemes). Examples found in the sample are below:

(64) *buffer, sewage, eigenvalue, hind-sight*

Buffer "a person or thing that lessens shock or protects from damaging impact, circumstances, etc." is according to the OED not related to *buff* "polish" but to a different lexeme *buff* "to act and sound as a soft inflated substance does when struck, or as the body does which strikes it", now obsolete. Therefore, *buff* in *buffer* is considered to be a monocollocable component. *Sewage* is similar in containing *sew* which resembles a common English verb, but to which it is not related. It relates to *sewer* and has according to the OED been formed from it, but there is no other lexeme in which *sew* is used in this sense. It is debatable whether this is a correct analysis from a purely synchronic perspective, as the two components could also be analysed as instances of the existing morphemes *buff* and *sew* with a completely unpredictable meaning. However, it makes more sense to keep the two unrelated lexemes distinguished from each other. *Eigenvalue* is a mathematical term and an instance of loan translation with one part being kept in the original form. Unassimilated borrowings are one of the possible sources of monocollocable components. *Hind-sight* or "the ability to understand, after something has happened, what should have been done or what caused the event" is a

borderline case because *hind* is used as a free morpheme meaning "rear", but it mostly collocates only with parts of the body of animals (*hind legs*). It also occurs as part of *behind* (which may be considered simple from the synchronic point of view) and several other rare words such as *hindwards*.

6.2.4.3. Semantic incompatibility

Semantic incompatibility is a type of collocational anomaly which is only secondary to semantic anomaly of components described in 6.2.3., as only those semantically incompatible components which are not used in their regular meaning can be used together in a meaningful lexeme.

There are probably at least two stages of semantic incompatibility. The first stage includes incompatibility of a lower degree. These lexemes are typical in that their parts do not make sense together in the word-formation meaning. The category is exemplified by these cases:

(65) *interviewer, highlight, chairperson, double-blind*

As far as *interviewer* is concerned, it may be difficult to see how the two parts (*inter-*, "between, mutually", and *view*, "see") relate to each other already in its base, *interview.* In *interviewer*, there is moreover the agentive suffix *-er* which seems to be incompatible with the reciprocity sense of *inter-*. *Highlight* is formed from compatible parts if it is used as a noun, but if it is used as a verb "to bring notice or emphasis to" the relation of *high* to *light* is hard to define (this might be, however, solved by approaching converted lexemes as instances of relisting in which the original word-formation structure is backgrounded, and the converted lexeme is treated as a simple form). *Chairperson* "a person who presides over a company's board of directors, a committee, a debate, an administrative department, etc" has also an opaque relation between *chair* and *person* in the word-formation meaning (due to metonymy). *Double-blind* "of or relating to an experiment to discover reactions to certain commodities, drugs, etc, in which neither the experimenters nor the subjects know the particulars of the test items during the experiments" is opaque because a quality such as *blind* cannot be *doubled* in the common sense of *blind*. It can be seen from the adduced examples that except for the word *interviewer*, which has a borrowed base incidentally containing components used also as morphemes in English, the rest of the examples discussed above are based on metaphor or

metonymy. This indicates that it is usually some semantic shift that is the source of semantic incompatibility.

The second, and more prototypical, type of semantic incompatibility is complete incompatibility of senses. This concerns especially semantic contrast between a modifier expressing a quality and a head which cannot be the bearer of the quality. This is exemplified by the following lexemes:

(66) *deadline, dead end, fast food, interface, airport, bed-rock, soap opera*

Deadline "a time limit" and *dead end* "cul-de-sac" or "a situation in which further progress is impossible" are instances of animate-inanimate semantic clash. *Fast food* "food that requires little preparation before being served" is a clash between moving and stable, *interface* has components incompatible in position (*inter-* means in between, whereas *face* is the surface of something). *Airport* „a place where aircraft land and take off", *bed-rock* "solid rock beneath the soil" and *soap opera* "a serialized drama, usually dealing with domestic themes (originally sponsored by soap manufacturers)" all consist of two components from different lexical fields, which is the source of semantic clash and opaqueness of the word-formation meaning.

6.2.4.4. Tautology

A specific case of collocational anomaly, which can be seen as having both formal and semantic aspects, is tautology. Tautology in word-formation is sometimes also called redundancy or pleonasm although the terms refer traditionally to different concepts (cf. Szymanek, 2015). Tautology can be of two kinds, although they often overlap. The first type is related to form as exemplified below:

(67) *packaging, dosage*

(68) *capacitance, fractionation*

Examples in (67) are anomalous in containing a formal marker which does not change (or at least not significantly) the meaning of the base: *packaging* "the container or covering that something is sold in" is very similar in meaning to package "a small container in which a quantity of something is sold" (the base has several other meanings in addition). The definitions taken from CD show that the meaning is basically the same. Similarly, *dosage* is

defined as “the amount of a medicine or drug that someone takes or should take” and its base, *dose,* as “a measured amount of it which is intended to be taken at one time”.

Examples in (68) are terms from specialized fields and unlike examples in (67), they carry a specific meaning, different from the base. What is anomalous here is the combination of two affixes which are normally of a very similar function: *-ity* + *-ance* in *capacitance* “the property of a system that enables it to store electric charge” and *-ion* + *-ation* in *fractionation* “the different condensation from a mixture of vapours in different parts of a separator or reactor”

The second type of word-formation tautology found in the sample is semantic tautology, where the meaning of the modifier is already included in the meaning of the head:

(69) *shot-gun, machine gun, driveway, problem-solve*

(70) *upsurge, higher-up, age-old*

Examples in (69), *shot-gun* “a shoulder firearm with unrifled bore designed for the discharge of small shot at short range and used mainly for hunting small game”, *machine gun* “a rapid-firing automatic gun, usually mounted, from which small-arms ammunition is discharged” and *driveway* “a private road for vehicles, often connecting a house or garage with a public road” are instances of meaning inclusion, which, however, differ in their lexical meaning from their head because they are used as terms involving an opaque specification of meaning. *Problem-solve* “to find solutions to problems, esp. by using a scientific or analytical approach” (a back-formation from *problem-solving*), has a redundant component *problem* since *problem* is a prototypical object of *solving*. The lexical meaning is also specialized.

Lexemes in (70) are instances similar to (69) in being made of two parts which are very similar in meaning: *upsurge* “a rapid rise or swell” is formed by compounding *surge* “a sudden increase” (one of several senses) and “up”, which is semantically included in the meaning of *surge. Higher-up* “a person of higher rank or position” includes in its word-formation meaning two semantically very close components. The same applies to *age-old* “very old”, where the noun *age* is a modifier which serves as an intensifier.

6.2.4.5. Formal and collocational anomaly – quantitative data

The following table provides quantification of the discussed types of formal or collocational anomaly:

CATEGORY	SUBCATEGORY	NUMBER OF LEXEMES	% IN OED SAMPLE
formal anomaly	**anomalous component form**	23	4.6
	anomalous suffix	9	1.8
	other formal anomalies	16	3.2
formal and collocational anomaly	**formally anomalous combination**	25	5.0
	tautology	14	2.8
	semantic incompatibility	33	6.6
collocational anomaly	**low collocability**	5	1.0

Table 30: Formal and collocational anomaly in the OED sample

The main finding illustrated in Table 30 is probably that both these types of anomaly occur only occasionally in addition to semantic anomaly All in all, only 110 lexemes exhibit a formal or collocational anomaly[10]. forming just 22% of the OED sample, the rest is made up of lexemes primarily displaying some kind of semantic anomaly,

In conclusion of the preceding two sections dealing with different types of anomalies, it may be said that there are several axes along which the degree of idiomaticity can be measured: the **first axis** goes from transparent lexemes through lexemes having both transparent and opaque meaning to lexemes with only opaque meaning. The **second axis** goes from more systematic

[10] Collocational and formal anomaly often combine, and therefore the number of lexemes with any of these anomalies is lower than the sum of lexemes in Table 30.

meaning shifts, such as specialization of meaning or metonymy, through less systematic and more creative meaning shifts, such as metaphor or exocentricity, to most idiomatic instances of non-compositionality with an idiosyncratic relation between the word-formation and lexical meaning. The **last, third axis** goes from regularly formed lexemes with regularly combined components, through formal and collocational irregularities, to idiosyncratic instances of semantic incompatibility or extremely restricted collocability. It would therefore be simplifying to claim that the lexemes with a combination of more types of anomaly are all more idiomatic than lexemes displaying only semantic anomaly without formal irregularity, because the degree of idiomaticity can be very high along one axis only which will cause the lexeme to be perceived as highly idiomatic anyway.

6.2.5. Formal classification of lexical idioms

The present section summarizes the OED sample from the formal point of view, focusing on the word-formation types. The distribution of word-formation processes in the sample is illustrated above in Table 24 and Figure 8. Specific features of each word-formation type of lexical idioms are discussed below. The examples are generally only listed in this section as most have already been discussed in the preceding sections.

6.2.5.1. Derivatives

The sample includes 158 derivatives (31.6% of the OED sample lexemes) including 98 nouns, 41 adjectives, 16 verbs and 3 adverbs. Especially the first two word-classes are represented sufficiently enough so that the categories can be studied in more detail.

If we attempt to analyse in detail the subclass of derivative nouns, we find that there are 73 instances of suffixation, 16 instances of prefixation and 9 instances of combined prefixation and suffixation. The range of affixes used is wide and they include all three onomasiological types according to Dokulil's classification (cf. 3.4.1.). The mutational type is represented mainly by affixes *-er*, -or and *-ist*. Affix *-ist* (7 instances) is in all cases represented by lexemes with some kind of semantic shift, but regular form. On the other hand, suffixes *-er* and *-or* constitute a heterogeneous group of formally regular instances of meaning shift (*processor, reactor, organizer*), formally anomalous lexemes (*steamer, grader, tanker,*

capacitor) and an instance with a monocollocable component (*buffer*). The mutational type is also represented by exocentric formations with various affixes (*detective, documentary, mutant, wireless, microwave*). Instances of transpositional onomasiological type are also very common (approx. 30 % of noun derivatives). They are mostly idiomatic due to specialization of meaning (*coverage, connectivity, mobilization, activism*) or another shift (*dynamism, magnetization, shrinkage*), but there are also instances which represent less systematic categories of idioms with formal anomalies (*dosage, sewage, flotation, capacitance*). The modificational type is also represented in the sample, especially by prefixation (15 instances), but also by diminutive affixes. Prefixation is in almost all cases represented by Latinate prefixes and the opaqueness is caused by meaning specialization of the lexeme, mostly in field-specific language (*interface, infrastructure, subtitle, subroutine, subway*).

The fact that Dokulil (1978) concludes that it is especially the mutational category which exhibits discrepancy between word-formation and lexical meaning and our sample shows that discrepancies are quite frequent also in the other two types can be explained as follows. My criteria of discrepancy are broader, including also semantic shifts (especially meaning specialization, metaphor and metonymy) and formal and collocational anomalies which are not dealt with in Dokulil.

The subclass of derivative adjectives contains almost only instances of semantic anomaly (there are only 4 instances with formal or collocational anomaly). Most examples are instances of one of the semantic subtypes (metaphor: *viral, colourful, multidimensional, telling*; specification: *nuclear, distal, inter-war, institutionalized*; amelioration: *exceptional, resourceful*). From the formal point of view, the category is more homogeneous, with four important suffixes (*-ic, -al, -ing, -ed*), several cases of prefixation (which are very similar to prefixed nouns discussed above, e.g. *infra-red, multimedia, inter-war*).

6.2.5.2. Compounds

Compounds represent the largest group in the OED sample with 283 instances (56.6%). The most numerous group of compounds are nouns (229 instances). The analysis will first focus on noun compounds excluding particle compounds. They will be discussed separately as they form a distinct, somewhat problematic, group from the point of view of phraseology.

There are 133 noun compounds excluding particle compounds. From the formal point of view, noun compounds are mostly represented by the structural type N-N (73 instances), ADJ-N (25 cases) and N-V (8 cases). The sample includes also occasional instances of other types, such as V-N, neoclassical formations, ADV-N, PREP-N, N-ADV and phrasal compounds.

Most of N-N compounds are anomalous only semantically. Ten cases contain some collocational anomaly (mostly semantic incompatibility due to semantic shift: *soap opera, airfield*), but the rest is regular from the point of view of form and collocability. Metaphor (*viewpoint, horsepower, headlight*), metonymy (*skyline, airforce, manpower*) and meaning specialization (*wave length, work station, artwork*) are the most common subtypes of anomaly causes.

ADJ-N compounds are less frequent, but the subclass is very similar in the analysed aspects: there are no formally anomalous lexemes in this category and only 4 instances of collocational anomaly (semantic incompatibility due to semantic shift). As far as semantic subtypes are concerned, there are instances of metaphor (*hot spot, dead end*), metonymy (*fast food, bad news*), specification (*natural gas, open system*) and exocentric lexemes (*high-rise, heavyweight, freelance).*

There are 40 adjective compounds in the sample and the class is very heterogeneous, involving instances of meaning specialization (part-time, high-level), metaphor (double-blind, undercover), metonymy (white-collar, on-line), instances of idiosyncratic meaning discrepancy (hung-up, way back, upfront) and instances of formal and collocational anomaly (one-one, overall, overseas, age-old).

Particle compounds are represented by 96 instances in the OED sample. All particle compounds identified during the collection of the sample were included in the sample as they are all more of less semantically opaque. This is especially because of the semantic vagueness of their components – particle compounds are based on phrasal verbs and this vagueness is typical of them. It has been mentioned above that particle compounds are problematic from the point of view of phraseology. Their unclear status is caused by the clash between two aspects: the first aspect is their discrepancy between word-formation meaning and lexical

meaning: e.g. *takeover* has the lexical meaning "the act or an instance of assuming control or possession", but its word-formation meaning consists of a combination of the extremely vague verbal component *take* and the particle component *over*, which is at least as vague as *take*. The discrepancy between the indistinct word-formation meaning and the highly specific lexical meaning is self-evident, and the particle compound should be therefore seen as a lexical idiom.

On the other hand, there is the second aspect of its formation: the particle compound is formed directly and regularly from the phrasal verb *take over* "to assume the control or management of". The formation of particle compounds from phrasal verbs is very productive and systematic. Whether we should see the particle compounds as idiomatic or regular, depends on our perspective and aim of study. If phraseology is seen as the study of combinatorial processes, then particle compounds will be probably seen as regular as the word-formation process of forming particle compounds from phrasal verbs is highly systematic and regular. In contrast, if phraseology is understood as the study of combinations of components in the language, then particle compounds must be seen as idiomatic.

From the formal point of view, there are two types of particle compounds: the regular one, represented by the type V+P, e.g. *set-up, break-down, make-up*. The verb-to-noun conversion is in this case accompanied by stress-shift to the first syllable. The second, less common type is P+V, e.g. *outcrop, output, upkeep*. It is not clear how to describe these formations as of the 12 instances in the sample, 7 of them are according to the OED attested earlier than the corresponding complex verbs of the type P+V: *output, uptake,* etc. We may therefore assume that the noun is formed not from phrasal verbs but from these complex verbs. On the other hand, the OED also mentions cases where the noun is attested earlier than the verb: *outcrop, upkeep, upgrade* or simultaneously with the verb (*bypass*) or only the noun is listed in the OED (*throughput*). It seems therefore that this structural type is heterogeneous in its origin, but a more extensive analysis is needed to reach a definite conclusion.

Focusing on the former and more common type of particle compounds, we may see that as far as its semantic compositionality is concerned, the degree of idiomaticity is not the same for all members of this group. This has been partly described in section 6.2.3. where some particle

compounds were included in the subtype of meaning specialization and some of them were described as subtypes of the unspecified category. This distinction is based on the relation between the word-formation meaning and lexical meaning, but there are also differences with respect to the phrasal-verb meaning.

Some particle compounds correspond closely to the meaning of the relevant phrasal verb, e.g. *set-up, break-down, come-back, lay-out, drop-out*. This group will be seen as less idiomatic, and this concerns especially those lexemes which are based on phrasal verbs with relatively high specification (in comparison with typical phrasal verbs): *wash-out, trade-off, clean-up, speed-up*. As these latter instances are regularly formed and the meaning is quite transparent, they are least idiomatic of the whole class.

The second type includes particle compounds whose meaning is closely related to the meaning of the phrasal verb, but with a restricted set of senses, which is a common phenomenon associated with conversion. Examples include *call-up* (where the most common verbal meaning "to telephone" is not listed in the nominal senses) and *cover-up* (which lacks the concrete, literal, meaning of the corresponding phrasal verb, meaning only "concealment or attempted concealment of a mistake, crime, etc").

The third type includes particle compounds with a meaning derived from the meaning of the corresponding phrasal verb, but with meaning specialization. Instances of this type include: *get-out* "an escape from a difficult situation", *show-up* "a police identification parade" and *slow-down* "a protest in which workers deliberately work slowly and cause problems for their employers".

The most idiomatic cases exhibit some additional discrepancy between the meaning of the phrasal verb and the nominal compound. An instance of this type is *set-aside* "a scheme in which a proportion of farmland is taken out of production in order to reduce surpluses or maintain or increase prices of a specific crop", *pick-up* "a pickup track", *go-round* "one of a series of actions, encounters, meetings, etc., often one involving a conflict or fight" and *spin-off* "any product or development derived incidentally from the application of existing knowledge or enterprise".

6.2.5.3. Combined formations

The category of combined formations includes 45 lexemes in the OED sample. The category contains all lexemes which consist of at least three distinct components and both the process of derivation and composition are involved in word-formation. A large number of these formations, however, are instances of two subsequent processes which do not interfere much with each other - neither formally, nor semantically. Nevertheless, they are different from two-component structures because they contain at least one more meaningful component.

The most interesting instances of combined formations are synthetic compounds, which are defined as compounds including a verbal component and its arguments. Typical examples are *bus driver* or *bookseller*. It has been already mentioned that due to its onomasiological structure, the type is usually transparent. However, in spite of this, some instances were also included in the sample as lexical idioms due to some kind of semantic shift. They include *care-taker* (meaning specification), *pacemaker* (metaphor), *bread-winner* (metonymy + semantically anomalous component), *loudspeaker* (semantically anomalous components), *bestseller* (meaning specialization and anomalous non-agentive referent).

The second group similar to synthetic compounds are adjectives of the type *blue-eyed* which are special by the simultaneous process of compounding and affixation. The sample includes among others adjectives *open-ended* (metaphor) and *two-dimensional* (metaphor).

Other instances of combined formations include deverbal compound nouns such as *fundraising*, *word processing* or *self-defeating*, particle compounds with affixation such as *ongoing, upheaval or overriding*, compound adjectives with a deverbal component such as *far-reaching, cross-sectional*.

In summary, it seems that combined formations should be seen as a structural type of lexical idioms along with derivatives and compounds although they are not typical representatives of the lexical idiom category. It also seems useful not to draw a strict line between simultaneous composition and affixation represented by synthetic or adjectival compounds and subsequent application of composition and affixation, as both may contain the same number of lexical components and may be therefore comparable with respect to their descriptiveness.

6.2.5.4. Other word-formation processes

It may be said that other word-formation processes, i.e. processes where components are not only morphemes or regularly formed bases, but also splinters, are generally idiomatic, as splinter-formations are always formally anomalous. The sample includes 14 instances of such formations, i.e. clipping compounds, affixed clipped components, a combination of both affixation and composition, and blending. Examples of these formations are in 6.2.4.1.

The last two sections of the present study will look at two subgroups of vocabulary which are, each for a different reason, associated with idiomaticity, i.e. terminology and pragmatic functions of lexical idioms.

6.2.6. Terminology and field-specific vocabulary

Although Čermák (2007a, cf. Table 3 in 2.9.) concludes that there are not many cases where terms overlap with idioms and that the main overlap is in the area of folk terminology based on metaphor (e.g. *kingfisher, foxglove*), the analysis of the sample has shown that in fact terms are similar to idioms in being often (more or less) non-compositional and, in addition, there may also be collocational or formal anomalies present. The non-compositionality does not concern prototypical systematically formed terms such as names of chemical substances (e.g. *sodium chloride, carbon dioxide*) but many other terms are formed partially arbitrarily (e.g. based on proper names, such as *camellia*, with heterogeneous bases, such as *neutron, electron, photon*, which can also be regarded as instances of collocational anomaly). A more systematic case of non-compositionality is that of meaning specialization. Meaning specialization is a logical consequence of the tendency towards formal economy which is common for both terms and non-terms, but which is very often significant for terms because of their highly specified reference which cannot be covered fully by a single lexeme. An instance of this phenomenon would be *resistor* with broader word-formation meaning ("somebody or something that resists") and specialized lexical meaning ("a device which is designed to increase the ability of an electric circuit to stop the flow of an electric current through it"). Semantic specialization is further discussed in connection with general vocabulary in 6.2.3.1. Apart from semantic anomalies, there may also occur instances of

formal (and collocational) anomalies, such as the Latinate-vernacular combination *infrared*, an instance of medial clipping in *rotor* or an anomalous combination of a noun and suffix *-or* in *capacitor*.

Terminology certainly must be distinguished from general language (i.e. the proper scope of phraseology), but it is rather difficult to decide where exactly the borderline between terminology and general language is. Apart from natural sciences terminology, such as names of substances and medical terms, which are typically formed with Latinate bases, neoclassical combining forms and Latinate or field-specific affixes, there are terms using general language bases, such as *capacitor* and *processor* from the area of technology, but also *value-added* from finance or *feminism, capitalism* from social sciences. And, in addition, speakers may perceive differences between *socialist*, which can intuitively be more likely classified as a term, and *leftist*, which is close to general language (also due to its native base and more vague meaning). Nevertheless, all these cases of marginal terms could be also described in terms of phraseology as non-compositional lexemes with meaning specialization of their lexical meaning.

We do not attempt to draw a clear dividing line between these categories in the present study as this would require a separate analysis focused on terms and terminology. Nevertheless, the sample does not include Latinate and neoclassical terms from natural sciences; social sciences terminology is also excluded due to its abstractness which is problematic when comparing the word-formation and lexical meaning. On the other hand, terms with a concrete referent (mostly technical terms) and native components are included in the analysis and marked as field-specific vocabulary. Social concepts are included only if they have undergone (at least partly) determinization. Lexemes with a concrete referent but abstract, non-native, combining forms or opaque bases (such as *telephone, telegram*) were not included.

There are 93 lexemes in the sample marked as terminology and field-specific language even after excluding the most central terms by the rules described above. The high frequency of terms and field-specific lexemes is a drawback of the selected methodology: since we aimed at filtering out old and fossilized formations, we have decided to search among lexemes which have entered the OED since 1800. However, this period is also a period of extensive

development of science and humanities and therefore the basis for the sample included a disproportionally high number of terms and other field-specific lexemes.

The following tables summarize the distribution of word-classes and word-formation processes among terms and field-specific lexemes included in the sample:

TERMINOLOGY AND FIELD-SPECIFIC VOCABULARY	
WORD-CLASS	**NUMBER OF LEXEMES**
V	2
ADJ	11
N	81
TOTAL	94

Table 31: Distribution of word-classes in terms and field-specific vocabulary in the OED sample

TERMINOLOGY AND FIELD-SPECIFIC VOCABULARY	
WORD-FORMATION PROCESS	**NUMBER OF LEXEMES**
COMPOUNDS	42
DERIVATIVES	41
COMBINED FORMATIONS	6
OTHER	5
TOTAL	94

Table 32: Distribution of word-formation processes in terms and field-specific vocabulary in the OED sample

Word-class distribution is similar to the corresponding numbers in the whole OED sample, although the proportion of nouns is even higher than in the whole sample. This is certainly

due to the typical function of terms, which is naming concepts, inventions, etc., which is a function normally associated with nouns.

The distribution of word-formation processes in this subgroup does not deviate much from the whole OED sample: compounds are most frequent (*home page, wavelength, mass media*), derivatives come second (*distal, sensor, systemic, globalization*), and there are a few combined formations (*networking, value-added, data processing, shareholder*) and a few other types (clipping compound: *redox*, clipping derivative: *soccer*, blend: *catalyse*).

Most lexemes in this group have idiomatic meaning only (70 cases). Several examples of this type are presented in example set (71):

(71) *shareholder, labour force, distal, natural gas, ultrasound, set-aside* (n.)

Several lexemes however have both idiomatic and non-idiomatic meaning (72):

(72) *nuclear, processor, fluorescent, preterm, open system*

As far as the semantic subtype is concerned, most lexemes of this group are marked as cases of meaning specialization (49 lexemes). This is not surprising, and we have explained that this kind of shift is rather problematic in terms of phraseology, as it is not really anomalous on the level of words. Examples of this type are listed below in (73):

(73) *processor, privatization, by-pass* (n.), *spreadsheet*, *leftist* (n.), *write-down* (n.)

There are also sporadic cases of other semantic subtypes within the category of terminology and field-specific vocabulary, 5 instances of exocentric formations (nouns: *adrenal*, *deductible*), 5 instances of metaphor (*website*, *dumping, horse power*), 3 instances of metonymy (*workforce, labour force*) and 2 instances of tautology (*capacitance, shot-gun*). Some of these lexemes were already discussed in the sections on the respective categories.

6.2.7. Pragmatic functions of lexical idioms

The pragmatic function of idioms is discussed by Čermák (2007a: 91-93), who is especially focusing on the traditional, collocational and propositional, idioms. The three aspects discussed by him are expressive, symbolic and evaluative aspects of meaning.

The expressive aspect of meaning is that part of connotative meaning which expresses some attitude of the speaker towards the information conveyed. Examples of English collocational and propositional idioms with the expressive function are *a hot potato* or *to be barking up the wrong tree*. These expressions are marked as informal and they have an equivalent expression in English which is neutral. The expressive connotation was not marked systematically in the sample, but it seems that it is much more marginal as lexical idioms are typically a part of neutral vocabulary. Instances of expressive lexical idioms can be nevertheless found in certain subgroups of lexical idioms: particle compounds are often informal (which is a feature inherited from the corresponding phrasal verbs). In addition, phrasal and exocentric compounds can be also expressive. Examples of expressive particle compounds are in (74), examples of expressive phrasal compounds in (75) and exocentric compounds in (76):

(74) *standout, sort-out* (both also exocentric)

(75) *must-be, no-good*

(76) *higher-up*, *know-nothing* (also phrasal)

Nevertheless, it must be concluded that even in the subclasses illustrated above, the expressive connotation is not common and that the vast majority of examples in the sample are more or less neutral in this aspect and even if they are marked as expressive, the intensity seems lower than in collocational and propositional idioms.

The second pragmatic aspect discussed by Čermák is the symbolic component of meaning which is associated with metaphor or metonymy. In short, the connotative aspects are transferred from the vehicle to the tenor. The symbolic component is a typical feature also in lexical idioms based on metaphor or metonymy which are discussed above in 6.2.3.2. and 6.2.3.3.

The third aspect is the evaluative component. As Čermák explains, the evaluative component adds connotative meaning on the scale good-bad. The sample was systematically analysed in this respect and it contains 44 lexemes (8.8 %) of lexemes of this type, which is not a small number, but it must be noted that the class of evaluative items was defined here very broadly, containing also not strictly evaluative lexemes, but also instances less central where the evaluative function is a component of a semantic shift (ameliorative, deteriorative,

euphemistic, pejorative aspect). In addition, the data are not compared with a corresponding sample of collocational and propositional idioms to see of the proportion of about 9 % is corresponding to the proportion of evaluative idioms at other structural levels. Examples of positive connotation are in (77), examples of negative connotation in (78)

(77) *exceptional, standout, must-see, phenomenal, knowledgeable*

(78) *no-good, drop-out, mechanistic, one-dimensional, no-account*

If I were to attempt to draw a conclusion as to these findings, it seems that both symbolic and evaluative components of meaning are represented to a significant extent in the class of lexical idioms, but the expressive component is very marginal in comparison to collocational and propositional idioms. Especially because of the lack of expressive function, which appears to be very important in other types of idioms, lexical idioms seem to be different from collocational and propositional idioms in their pragmatic function.

7. Conclusions

The aim of the dissertation is to establish the category of lexical idioms in English by means of a two-stage analysis of vocabulary samples. The theoretical part introduced the discipline of phraseology (Chapter 2) and then summarized references in the literature to idiomaticity on the lexical level (Chapter 3), arguing that although phraseologists consider it canonical to define phraseology as dealing with polylexical units, the linguistic literature in fact does not lack references to idiomatic derivatives and (especially) idiomatic compounds. Finally, the theoretical part introduces Čermák's account of lexical idioms as part of the phraseological study. His concept of lexical idioms is also the main theoretical source for the present study (Chapter 4).

The empirical part outlines the objectives and research questions and provides a provisional definition of lexical idioms and by description of the data used in the analysis (Chapter 5). To begin with, lexical idioms are defined as single-word lexemes formed as combinations of components which are anomalous semantically and/or collocationally and/or grammatically. The analysis itself (Chapter 6) is divided into two stages. The aim of the first stage was to analyse randomly-chosen English single-word lexemes with respect to their idiomaticity and adapt the provisional definition of lexical idioms to the specific situation of the English lexicon. On the basis of the findings a new definition was formulated and used to retrieve a new sample of items meeting the criteria for lexical idioms. This second sample was then analysed in detail in the second stage of the analysis to identify different types of English lexical idioms and the source of their idiomaticity.

The first sample consisted of 1 000 randomized lexemes retrieved from the BNC. The lexemes were classified into several categories to find out which irregularities should be included among the criteria for lexical idioms. The sample included 319 simple lexemes and 681 complex lexemes. The latter category was then classified in terms of formal, collocational and semantic regularity or anomaly. Of the 681 complex lexemes 381 were assigned to the category of regular formations and the remaining 300 lexemes displayed at least one of the three types of anomaly (altogether the lexemes showed 407 anomalies): formal anomaly

occurred in 158 lexemes, collocational anomaly in 81 lexemes and semantic anomaly in 168 lexemes.

It became clear during this first stage of analysis that in order to identify lexical idioms in English, it was necessary to define more precisely what exactly is meant by an anomaly on the lexical level. The BNC sample analysis showed that transferring directly the criteria for Czech lexical idioms chosen by Klötzerová (1997) is not plausible for English because this method retrieves more than 50% of all complex lexemes as formally anomalous mainly because of unproductive word-formation processes and because of a high number of assimilated borrowings of Latin or Greek origin. I have therefore decided to consider regularity and anomaly in terms of existing analogies in the system and not in terms of current productivity. This strategy is based on the finding that the combination of typical characteristics of established words (i.e. institutionalization and lexicalization) and the peculiarities of the English lexicon (a high proportion of Latinate formations and borrowings) would have the following effect: if current productivity were taken as the main formal criterion of lexical idioms, the proportion of idioms in vocabulary would be too high, not corresponding at least roughly to the understanding of idioms as exceptions from the norm and in addition, a large proportion of such idioms (41 %) in the BNC sample would be semantically transparent, which is not typical of idioms either.

The BNC sample has also shown that semantically anomalous lexemes are the only group which is structurally different from the whole class of complex lexemes. In particular, in the whole class of complex lexemes, the category of derivatives was much more common (76.8 %) than the category of compounds (17.5 %), but for lexemes with idiomatic meaning, the proportion was more in favour of compounds (47.6 % of compounds and 46.9 % of derivatives). A similar difference was spotted in the analysis of word-formation processes: while the whole class of complex lexemes in the sample includes 47.3 % of nouns and 38.8 % of adjectives, lexemes with idiomatic meaning include 67.2 % of nouns and 20.3 % of adjectives. Similar differences did not occur with any other type of anomaly.

The problematic subclasses of Latinate borrowings and Latin-based formations, the specific word-class and word-formation process distribution among semantically anomalous lexemes

and formal features typical for all words (formal fixedness, lower degree of productivity of word-formation in comparison with syntax) have led to the decision to consider semantic anomaly to be the primary criterion in defining lexical idioms and regard formal and collocational anomalies as secondary criteria which serve to further raise the degree of idiomaticity of semantically non-compositional lexemes.

Thus the first stage of the analysis made on the BNC sample provided the answer to research question number 2: semantic anomaly is considered primary in lexical idioms and both formal and collocational anomaly are still regarded important, but their effect is limited, as some subclasses of formal anomaly typical of polylexical idioms are not relevant for lexical idioms and productivity has been replaced by analogy.

The amended definition was then applied when collecting the OED sample of 500 lexical idioms for the second analysis. In this second stage, the selection was restricted to lexemes formed within English with the first entry in the OED after 1800 to exclude most instances of Latin and Greek borrowings and early Latin-based formations within English. This analysis focused primarily on semantic anomalies, analysing above all the possible types of discrepancy between word-formation semantics and the lexical meaning of the components. Nevertheless, formal and collocational anomalies were studied as well, and the data also showed several subtypes of these anomalies. In addition, the correlation between formal structure (word-class and word-formation type) and semantics (the subtype of anomaly) was inspected in a separate section. The final sections briefly discussed the relation of lexical idioms to terminology and the pragmatic functions of lexical idioms.

The findings based on the OED sample revealed that the range of possible semantic anomalies is quite wide (cf. research question number 6). Two types of semantic shift in the meaning of the components are commonly described in the literature in connection with idiomaticity, and they occurred also in my sample: metaphor (21 % of the sample) and metonymy (7 %). In addition, the sample included many instances of meaning specialization (24.6 %), relatively many cases of exocentricity (non-hyponymous lexemes; 10%), but only few examples of meaning generalization (2%), and occasional instances of other discrepancies and shifts such as hyperbole, determinization, amelioration, deterioration and euphemistic and pejorative

expressions. As far as exocentricity is concerned, the study discussed not only exocentric compounds which are traditionally described in the literature, but the term was also extended to derivatives formed by ellipsis from multi-word expressions or by conversion. The data suggest that some semantic subtypes will probably lead to a higher degree of idiomaticity, while other subtypes of anomaly are only marginal in the category of lexical idioms. The two subtypes of semantic anomaly producing most idiomatic items seem to be metaphor and exocentricity. Metaphor is very idiomatic in that it is shifting not only the meaning of components but also their conceptual domain. Exocentricity is also highly idiomatic because of the formal discrepancy between the syntactic and the referential head. The least idiomatic types seem to be instances of change in connotation only (pejorative and euphemistic expressions), but also instances of meaning specialization. Specialization is to some extent present whenever a new word is coined, and there is a fuzzy borderline between the necessary meaning specialization due to formal limitations and the additional unexpected specialization.

Formal and collocational anomalies were quite rare. In fact, only 110 of the 500 lexemes exhibit formal and/or collocational anomaly. Nevertheless, several subtypes were recognized also in this group: lexemes with anomalous form of the component (e.g. a fossilized form or a splinter), derivatives with an anomalous suffix (a suffix signalizing a different word-class), phrasal compounds, lexemes with syntactic or morphological anomaly, lexemes with formally anomalous combinations of components, semantically incompatible components, with low collocability components and with components showing redundancy (both formal and semantic).

Several noteworthy findings were arrived at concerning the formal structure of lexical idioms. The category of lexical idioms in English divides into four main groups in terms of the word-formation process involved: compounds were the most common type (56.6 % of the OED sample), derivatives were second (31.6 % of the OED sample), and formations combining both composition and derivation were the third (8.8 % of the OED sample). The fourth group, splinter formations made up of splinters (i.e. fractions of words arising in blending) such as clipping compounds, blends, etc., was least numerous (3% of the OED sample). Hence these findings answer research question number 3: lexical idioms do include more than compounds and derivatives, and the other two groups, combined formations and splinter formations, have

not been described in the literature so far. Combined formations are probably more common in English than in Czech, which employs composition less often than English, and splinter formations are definitely typical of English (and rare in Czech). They may be problematic with respect to phraseology because once a splinter begins to be used in more formations, by analogy, it assumes the status of an affix and its use becomes systematic. Such formations then cease to be anomalous.

The results of the analysis also indicate which formal categories of lexemes prevail among lexical idioms (research question number 4). The word-classes most typical for lexical idioms are nouns and adjectives and the most prominent word-formation process is composition, although derivatives are certainly not negligible. The most central idioms (i.e. idioms with some kind of highly idiosyncratic anomaly, not a representative of any of the semantic subtypes described above) proved to be instances of compounds or derivative nouns, whereas derivative adjectives were mostly lexemes with both transparent and idiomatic meaning (based typically on metaphor or meaning specialization). A very distinct and numerous subclass (108 instances) is that of particle compounds. The research makes it clear that they must be included in the category of lexical idioms on account of their structure, but they may be regarded as regular in terms of the word-formation process. The analysis also indicated that this group is not homogeneous and includes instances with a varying degree of idiomaticity.

On the other hand, the sample shows that there are some structural types which tend to be transparent (cf. research question number 5). This concerns especially three suffixes which are highly productive and often fully transparent: adverbial *-ly* and adjectival *-ing* and *-ed*. This transparency can be no doubt attributed to their status which is close to that of inflectional affixes. More generally, it may be claimed, on the basis of the analysed data, that the proportion of transparent lexemes is higher among derivatives whereas compounds tend to be more often opaque. An interesting formal group with respect to phraseology is that of back-formations. Back-formed lexemes were not discussed very much in connection with either of the samples. The reason for this is that although back-formation is traditionally described as a minor word-formation process, the formations themselves must be seen as regular from the synchronic point of view if the basic criterion of regularity is analogy (actually back-formation is one of the major sources of compound verbs in contemporary

English). Finally, synthetic compounds also tend to be rather transparent, although the number of exceptions is higher in this case due to possible semantic shifts, especially metaphor, meaning specialization and metonymy.

The last section of the OED sample analysis deals briefly with pragmatic functions of lexical idioms. The data indicate that the symbolic component of the connotative meaning is typical of lexical idioms, especially of figurative ones and that evaluative connotation is not uncommon among lexical idioms (8.8 % in the OED sample). However, further research is needed to compare the results with corresponding data on multi-word idioms. By contrast, the expressive component, which is very typical of multi-word idioms, is only marginal among the lexical idioms of the OED sample, as they do not typically represent an alternative to a stylistically neutral expression, which is the case of collocational idioms. These data may also serve as a tentative answer to research question number 7 (expressivity as a feature of English lexical idioms).

The answer to the question of what determines the degree of idiomaticity displayed by a lexical idiom (cf. research question number 8), following from the sample analysis, is that there are at least three scalar axes or variables of idiomaticity. The first variable represents a scale from transparent non-idiomatic lexemes, through lexemes transparent, i.e. non-idiomatic, in some of their senses but opaque, idiomatic in others, to lexemes with only opaque idiomatic meaning. The second variable is a scale from full compositionality (no discrepancy between the lexical meaning and word-formation semantics), through less idiomatic discrepancies between the word-formation semantics and the lexical meaning such as changes in the connotation, meaning specialization and metonymy, to more idiomatic discrepancies such as metaphor and exocentricity down to major semantic discrepancies. Within this aspect, it was also demonstrated that there are differences even inside the semantic subtypes: if the discrepancy between the lexical and the word-formation meaning is based on some prototypical structure (e.g. a common conceptual metaphor), the lexeme is more transparent than if the discrepancy is purely idiosyncratic. The last variable is represented by

combinatorial regularities or anomalies in the formal structure. The scale starts with formally regular expressions from common morphemes and ends by formal and collocational anomalies where monocollocability is the extreme case.

axes/variables of idiomaticity		**degrees of idiomaticity properties**		
compositionality	**transparency/opacity**	transparent	transparent/opaque	only opaque
	discrepancy between lexical and word-formation meaning	no discrepancy	mild discrepancy due to systematic (e.g. metonymy) to less systematic, creative meaning shift (e.g. metaphor) in components	major, idiosyncratic discrepancy between lexical meaning of component and WF semantics
combinatorial regularity / anomaly		formal/collocational regularity	formal and/or collocational irregularity	major combinatorial idiosyncrasy, anomaly

Figure 13: Axes and degrees of idiomaticity

Finally, we can return to the most general and theoretical question: is it meaningful to study anomalous combinatorial relations below the level of the word within and by means of phraseology? Judging by the results of the analysis presented in this study, it is justifiable to say that components combined below the level of the word behave to a certain degree similarly to elements of phraseological units above the lexical level. The most important similarity between the two levels resides in the fact that combinations on both levels are normally based on a set of rules which can be sometimes disrupted by anomalies. This is essentially the main argument for the inclusion of lexical idioms into phraseology.

Nevertheless, there are also several dissimilarities which are so significant that they cannot be disregarded. The most important one is the difference between word-formation and syntactic combinability. In particular, rules for combining of words are much more generally valid and therefore a potential anomaly is more distinct than is the case with sublexemic components, as their combinations are more restricted, especially by such limiting factors as institutionalization and lexicalization (and the subsequent phenomenon of blocking). Recent psycholinguistic research also shows that words are normally stored as one unit even if they

are formed regularly and they are decomposed only in special circumstances, which also differentiates them from (most) free combinations of words. In addition, the results show that there are additional differences between multi-word and single-word idioms in their pragmatic function which may point to their different nature. It is also typical that the most prominent group among lexical idioms is comprised of compounds, which are made up of two lexical bases and thus are nearest to the borderline between words and word combinations.

The findings of the study summarised above warrant the conclusion that it is indeed possible to study single-word lexemes in terms of phraseology but at the same time one must be aware of the differences and consider the whole category of lexical idioms as being on the periphery of the domain of idioms.

The present study has also defined certain problematic areas from the theoretical point of view which may be dealt with in further research. One of them is the actual aim of the study, the establishment of lexical idioms, and the implications arising from this aim. For instance, the definition of lexical idioms will differ in a theoretical structuralist study describing anomalous word-formation processes and in practical ELT application where the identified lexical idioms will be used for a vocabulary-teaching methodology. Another question only marginally discussed in this study is that of the origin of lexical idioms. It seems that anomalies are sometimes due to language development, sometimes to creative use of existing lexemes and sometimes they result from creative coinage of new lexemes. The matter certainly merits further analysis.

References

Ackema, P., & Neeleman, A. D. (2004). *Beyond Morphology; Interface Conditions on Word Formation: Studies in Theoretical Linguistics*. Oxford: Oxford University Press.

Aitchison, J. (2012). *Words in the Mind: An Introduction to the Mental Lexicon*. Blackwell: Wiley-Blackwell.

Allerton, D. J. (1982). *Valency and the English verb*. London-New York: Academic Press.

Aronoff, M., & Fudeman, K. (2011). *What is morphology?* Malden-Oxford: Wiley-Blackwell.

Barkema, H. (1996). Idiomaticity and terminology: A multi-dimensional descriptive model. *Studia Linguistica* 50(2), 125-160.

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

Bauer, L. (1998). When is a sequence of two nouns a compound in English? *English Language & Linguistics* 2(1), 65-86.

Bauer, L. & Huddleston, R. D. (2002). Lexical word-formation. In Huddleston, R. D. & Pullum, G. K. *The Cambridge Grammar of English Language*. Cambridge: Cambridge University Press, 1621-1721.

Bauer, L., Lieber, R., & Plag, I. (2013). *The Oxford reference guide to English morphology*. Oxford: Oxford University Press.

Bauer, L. (2016). Re-evaluating exocentricity in word-formation. In D. Siddiqi & H. Harley (eds.), *Morphological Metatheory*. Amsterdam/Philadelphia: Benjamins, 461-477.

Bauke, L. S. (2017). *A content matching analysis for compounds and idioms*. Roots V. Roots V, Queen Mary University London (and UCL) presentation. http://www.crissp.be/wp-content/uploads/2015/04/bcgl8-Bauke.pdf

Beard, R. (1977). On the extent and nature of irregularity in the lexicon. *Lingua* 42(4), 305-341.

Beard, R. (1987). Lexical stock expansion. In E. Gussmann (ed.), *Rules and the Lexicon: Studies in Word Formation*. Lublin: Redakcja Wydawnictw KUL, 23-41.

Beck, D., & Mel'čuk, I. (2011). Morphological phrasemes and Totonacan verbal morphology. *Linguistics* 49(1), 175-228.

Benczes, R. (2005). Metaphor-and metonymy-based compounds in English: a cognitive linguistic approach. *Acta Linguistica Hungarica* 52(2-3), 173-198.

Benczes, R. (2006). *Creative compounding in English: The semantics of metaphorical and metonymical noun-noun combinations.* Amsterdam-Philadelphia: John Benjamins Publishing.

Benczes, R. (2015). Are exocentric compounds really exocentric? SKASE Journal of Theoretical *Linguistics* 12(3), 54-73.

Benson, M., Benson, E., & Ilson, R. F. (1986). *Lexicographic description of English*. Amsterdam-Philadelphia: John Benjamins Publishing.

Biber, D. et al. (1999) *Longman grammar of spoken and written English*. Edinburgh: Pearson Education Ltd.

Bloomfield, L. (1965). *Language history*; from Language (1933 ed.). New York: Holt, Rinehart and Winston.

Bloomfield, L. (1984). Language (a new forward by C. F Hockett). Chicago: University of Chicago Press.

Borgwaldt, S., & Lüttenberg, D. (2010). *Semantic transparency of compound nouns in native and non-native speakers*. Poster presentation at 14th Morphological Meeting, Budapest, 13–16 May 2010..

Botha, R. P. (1968). *The Function of the Lexicon in Transformational Generative Grammar*. The Hague: Mouton.

Burger, H. (1998). *Phraseologie: Eine Einführung am Beispiel des Deutschen.* Berlin: Erich Schmidt.

Burger, H., Dobrovol'skij, D., Kühn, P., & Norrick, N. R. (2008). *Phraseologie / Phraseology*. Berlin-New York: De Gruyter.

Claiborne, R. (1990). *The roots of English: a reader's handbook of word origin*. Anchor Books.

Cowie, A. P. (1988). Stable and creative aspects of vocabulary use. In R. Carter & McCarthy, M. (eds.), *Vocabulary and language teaching*. London: Longman, 126-139.

Cowie, A. P. (1994). Phraseology. In R. E. Asher & Simpson J. M. (eds.), T*he Encyclopedia of Language and Linguistic*s. Oxford: Pergamon, 3168-3171.

Cowie, A. P. (1998). Introduction. In A. P. Cowie (ed.), *Phraseology: Theory, Analysis, and Applications*. Oxford: Oxford University Press.

Cowie, A. P. (2001). Speech formulae in English: problems of analysis and dictionary treatment. In G. van der Meer & ter Meulen, A. G. B. (eds.), *Making Senses: From Lexeme to Discourse. In Honour of Werner Abraham*, 1-12. Gröningen: Center for Language and Cognition.

Cowie, A. P. (2006). Phraseology In K. Brown (ed.), *Encyclopedia of Language & Linguistics* (Vol. 7). Oxford: Elsevier, 579-585.

Cruse, A. (2000). *Meaning in Language: An Introduction to Semantics and Pragmatics*. Oxford: Oxford University Press.

Čermák, F. (1982). *Idiomatika a frazeologie češtiny*. Praha: Karolinum.

Čermák, F. (2001). Substance of Idioms: Perennial Problems, Lack of Data or Theory? *International Journal of Lexicography* 14, 2001, 1–20.

Čermák, F. (2007a). F*razeologie a idiomatika česká a obecná. Czech and general phraseology*. Praha: Karolinum.

Čermák, F. (2007b). Idioms and Morphology. In H. Burger, D. Dobrovol'skij, P. Kühn, & N. R. Norrick (eds.), *Phraseology: An International Handbook of Contemporary Research*, 20-26. Berlin-New York: Mouton de Gruyter.

Čermák, F. (2017). Frazeologie a idiomatika. In P. Karlík, M. Nekula, J.Pleskalová (eds.), *CzechEncy - Nový encyklopedický slovník češtiny*. URL: https://www.czechency.org/slovnik/FRAZEOLOGIE A IDIOMATIKA (accessed 18 February 2018)

Dokulil, M. (1978). K otázce prediktability lexikálního významu slovotvorně motivovaného slova. *Slovo a Slovesnost* 39, 244-251.

Dokulil, M. (1986). III Tvoření slov. In J. Petr, et al. (ed.), *Mluvnice češtiny* (Vol. 1). Praha: Academia.

Downing, P. (1977). On the creation and use of English compound nouns. *Language* 53(4), 810-842.

Dressler, W. U. (2011). Grammar and pragmatics. In W. Winter (ed.), *On Languages and Language*: The Presidential Adresses of the 1991 Meeting of the Societas Linguistica Europaea. Berlin-New York: De Gruyter.

Ermakova, E. N., Zolnikova, N. N., Faizullina, G. C., Khasanova, M. S., & Khlyzova, T. N. (2015). Derivation and the Derivational Space in Phraseology as a Problem of the Language Contemporary Development. *Mediterranean Journal of Social Sciences*, MCSER Publishing, Rome-Italy, 6(3 S6), 335-340.

Filipec, J. et al. (1994). *Slovník spisovné češtiny pro školu a veřejnost*. Praha: Academia.

Giegerich, H. J. (2004). Compound or phrase? English noun-plus-noun constructions and the stress criterion. E*nglish Language & Linguistics* 8 (1), 1-24.

Gill, P. (2011). *Colouring Meaning. Collocation and connocation in figurative langauge*. Amsterdam-Philadelphia: John Benjamins Publishing Company.

Giora, R. (2002). Literal vs. figurative language: Different or equal? J*ournal of Pragmatics* 34. 487–506.

Gläser, R. (1986). *Phraseologie der englischen Sprache*. Leipzig-Tübingen: VEB Verlag Enzyklopädie/Niemeyer.

Gläser, R. (1988). The grading of idiomaticity as a presupposition for a taxonomy of idioms. In W. Hüllen & R. Schulze (eds.), *Understanding the Lexicon*. Tubingen: Niemeyer, 246–79.

Gläser, R. (1998). The stylistic potential of phraseological units in the light of genre analysis. In A. P. Cowie (ed.), *Phraseology: Theory, Analysis, and Applications*. Oxford: Oxford University Press.

Glinert, L. (2004). *The Grammar of Modern Hebrew*. Cambridge: Cambridge University Press.

Gottlieb, H (2012). Phraseology in the flux: Danish Anglicisms beneath the surface. In Furiassi, C., Pulcini, V., & González, F. R. (eds.), *The anglicization of European lexis*. Amsterdam-Philadelphia: John Benjamins*,* 169-198.

Granger, S. (2005). Pushing back the limits of phraseology: how far can we go? In Cosme, C., C. Gouverneur, F. Meunier & M. Paquot (eds.) *Proceedings of Phraseology 2005*. An Interdisciplinary Conference. Louvain-la-Neuve: Université catholique de Louvain, 165–168.

Granger, S., Paquot, M. (2008). Disentangling the phraseological web. In Granger, S., Meunier, F. (eds), *Phraseology. An interdisciplinary perspectiv*em. Amsterdam-Philadelphia: John Benjamins, 27-50.

Gries, S. T. (2008). Phraseology and linguistic theory: A brief survey. In S. Granger & F. Meunier (Eds.), *Phraseology: An interdisciplinary perspective*. Amsterdam-Philadelphia: John Benjamins Publishing, 3-25.

Haspelmath, M. (2002). *Understanding Morphology*. London: Hodder Arnold.

Haspelmath, M. & Sims, A. (2010). *Understanding Morphology*. London: Hodder Education.

Horn-Helf, B. (1997). *Kondensation als terminologisches Prinzip im Russischen*. Tübingen: Gunter Narr Verlag.

Howarth, P. (1998). Phraseology and second language proficiency. *Applied Linguistics* 19, 24-44.

Jackendoff, R. (2010). The parallel architecture and its place in cognitive science. In B. Heine & Narrog, H. (eds.), *The Oxford Handbook of Linguistic Analysis*. Oxford: Oxford University Press.

Kasik, R. (1997). Typology of Estonian and Finnish word-formation. The verb. In M. Erelt (ed.), *Estonian: Typological Studies II*. Tartu: Tartu University Press.

Kastovsky, D. (1982). *Wortbildung und Semantik*. Düsseldorf-Bern-München: Bagel/Francke.

Katamba, F., & Stonham, J. (2006). *Morphology* (2nd ed.). Houndsmills-New York: Palgrave Macmillan.

Kavka, S. (2009). Compounding and idiomatology. In R. Lieber & Štekauer, P. (eds.), *The Oxford Handbook of Compounding*. Oxford: Oxford University Press.

Klima, E. S., & Bellugi, U. (1979). *The signs of language*. Cambridge (Mass.)-London: Harvard University Press.

Klötzerová, P. (1997). *Lexikální frazémy a idiomy v češtině*. Diplomová práce. Praha: FF UK.

Klötzerová, P. (1998). Hranice frazeologie se posouvají. Lexikální frazémy v češtině. *Slovo a slovesnost* 59, 277–280.

Kooij, J. G. (1968). Compounds and idioms. *Lingua* 21, 250-268.

Körtvélyessy, L., Štekauer, P., & Zimmermann, J. (2015). Word-formation strategies: Semantic transparency vs. formal economy. In Bauer, L., Körtvélyessy, L., & Štekauer, P. (eds) *Semantics of Complex Words*. Cham-Heidelberg-New York-Dordrecht-London: Springer International Publishing, 85-113.

Kövecses, Z., & Radden, G. (1998). Metonymy: Developing a cognitive linguistic view. *Cognitive linguistics*, 9(1), 37-77.

Lakoff, G., & Johnson, M. (1980). *Metaphors We Live By*. Chicago: University of Chicago Press.

Levi, J. N. (1978). *The syntax and semantics of complex nominals*. New York, NY: Academic Press.

Li, C. N. (1971). *Semantics and the structure of compounds in Chinese*. PhD dissertation. Berkeley: University of California.

Lieber, R. (2009). *Introducing morphology*. Cambridge: Cambridge University Press.

Lipka, L. (2002). *English lexicology. Lexical structure, word semantics & word-formation*. Tübingen: Gunter Narr Verlag.

Lyons, J. (1968). *Introduction to Theoretical Linguistics*. Cambridge: Cambridge University Press.

McArthur, T. (ed.)(1992). *The Oxford Companion to the English Language*. Oxford: Oxford University Press.

Mel′čuk, I. A. (1976). *Das Wort* [Die Konversion als morphologisches Mittel]. München: W. Fink.

Mel′čuk, I. A. (1995a). *The Russian Language in the Meaning-Text Perspective*. Moskau-Wien: Wiener Slawistischer Almanach.

Mel'čuk, I. (1995b). Phrasemes in language and phraseology in linguistics. In M. Everaert, E. J. Van der Linden, A. Schenk (eds.), I*dioms: Structural and psychological perspectives*. Hillsdale, NJ: Lawrence Erlbaum Associates, 167-232.

Moon, R. (2015). Multi-word Items. In J. R. Taylor (ed.), *The Oxford Handbook of the Word*. Oxford: Oxford University Press.

Němec, I. (1968). *Vývojové postupy české slovní zásoby*. Praha: ČSAV.

Nerlich, B. (2006). Metonymy. In K. Brown (ed.), *Encyclopedia of Language & Linguistics* (Vol. 6). Oxford: Elsevier, 109-113.

Onysko, A., & Michel, S. (eds.)(2010). *Cognitive perspectives on word formation*. Berlin-New York: Walter de Gruyter Mouton.

Plag, I. (2003) *Word-formation in English*. Cambridge: Cambridge University Press.

Rodrigues, A., & Rio-Torto, G. (2013). Semantic coindexation: evidence from Portuguese derivation and compounding. In C. T. Pius ten Hacken (Ed.), *The Semantics of Word Formation and Lexicalization*,. Edinburgh: Edinburgh University Press 161-179.

Sinclair, J., P Hanks, et al.(1987) *Collins COBUILD English language dictionary*. London: William Collins Sons & Co. Ltd.

Sinclair, J. M. (1987). *Looking up: An account of the COBUILD project in lexical computing and the development of the Collins COBUILD English language dictionary*. London: Collins Cobuild.

Smirnitsky, A. I. (1998). *Leksikologija anglijskogo jazyka*. Moskva: MGU.

Szymanek, B. (2015). Remarks on Tautology in Word-Formation. In Bauer, L., Körtvélyessy, L., & Štekauer, P. (eds) *Semantics of Complex Words*. Cham-Heidelberg-New York-Dordrecht-London: Springer International Publishing, 143-161.

Štekauer, P. (1998). *An onomasiological theory of English word-formation*. Amsterdam-Philadelphia: John Benjamins Publishing Company.

Štekauer, P. (2005a). *Meaning Predictability in Word Formation: Novel, context-free naming units.* Amsterdam-Philadelphia: John Benjamins Publishing Company.

Štekauer, P. (2005b). Onomasiological approach to word-formation. In Štekauer, P. & Lieber, R. (eds.), *Handbook of Word-Formation*, 207-232. Dordrecht: Springer.

Tremblay, A., B. Derwing, G. Libben & C. Westbury (2011). Processing advantages of lexical bundles: Evidence from self-paced reading and sentence recall tasks. *Language Learning* 61 (2). 569–613.

Zhang, N. N. (2007). Root merger in Chinese compounds. *Studia Linguistica 61*(2), 170-184.

Sources

The British National Corpus (BNC) – Hoffmann, S., & Evert, S. (1996). BNCweb (CQP-edition). Accessed from bnc-web.lancs.ac.uk (last accessed on April 20, 2018).

The Collins English Dictionary online (CD) – accessed from www.collinsdictionary.com (last accessed on June 5, 2018).

The Oxford Advanced Learners' Dictionary online (OALD) – accessed from https://www.oxfordlearnersdictionaries.com (last accessed on June 5, 2018).

The Oxford English Dictionary (OED) – accessed from www.oed.com (last accessed on April 25, 2018).

Resumé

Tato studie se zabývá analýzou lexikálních idiomů v angličtině. Lexikální idiomy jsou definovány jako jednoslovné polymorfémní lexémy, které se vyznačují anomální kombinací komponentů. Základním cílem práce je zjistit, jestli je možné jednoslovné lexémy studovat, klasifikovat a popisovat v rámci frazeologie, tedy jako jednu z kategorií frazeologických jednotek. K odpovědi na tuto otázku se práce snaží dojít analýzou možných členů kategorie lexikálních idiomů a srovnáním jejich vlastností s vlastnostmi víceslovných frazeologických jednotek.

Studie navazuje na obecnou frazeologickou literaturu především kontinentálního a britského „frazeologického přístupu“, východisky jsou tedy například Burger (1998) a Cowie (1998), avšak hlavním zdrojem této studie je Čermák (2007a), který přináší myšlenku lexikálních idiomů v češtině. Čermák je jediným, který explicitně vztahuje pojem idiom i na jednoslovné jednotky a popisuje tedy lexikální idiomy jako jednu ze základních kategorií frazeologických jednotek. Jak je však prezentováno v kapitole 3 této studie, některé základní termíny frazeologie jsou v literatuře poměrně často používány ve spojitosti s jednoslovnými jednotkami. Toto se týká především popisu idiomatičnosti ve spojení s kompozity, avšak v literatuře je možné najít i odkazy na idiomatičnost slov odvozených. Na základě literatury je lexikální idiom definován jako jednoslovný lexém, který se vyznačuje anomálií sémantickou a/nebo formální a/nebo kolokační.

První část analýzy (kapitola 6.1.) je provedena na randomizovaném vzorku 1000 lexémů získaných z Britského národního korpusu (dále BNC vzorek). Vzorek je omezen na autosémantika (substantiva, adjektiva, verba a adverbia). Cílem této části analýzy je otestovat definici lexikálního idiomu a kritéria, která používá pro identifikaci lexikálních idiomů Klötzerová (1997, 1998). Za anomálie jsou považovány všechny odchylky od pravidelného produktivního tvoření slov. Vzorek obsahuje 319 slov monomorfémních, u kterých nelze o frazeologii hovořit, protože v nich nedochází ke kombinaci morfémů. 681 lexémů je zařazeno do třídy polymorfémních slov. Mezi oběma třídami je určitý překryv: mnoho výpůjček z klasických jazyků obsahuje v angličtině rozpoznatelný morfém, především afix, který je

však doplněn formou (bází) v angličtině nerozpoznatelnou a nesystematickou (např. *collide, defend, figment*). Tyto lexémy jsou považovány za monomorfémní a nejsou zařazeny mezi potenciální lexikální idiomy.

Třída polymorfémních slov je dále prozkoumána z hlediska výskytu možných anomálií. Ze 681 popisných lexémů je 381 lexému pravidelných sémanticky i formálně. Zbylých 300 lexémů vykazuje nějakou anomálii a vzorek celkem obsahuje 407 anomálií. Tyto lexémy jsou dále zařazeny do kategorií 1-4 podle konkrétního druhu anomálie. Pokud lexém vykazuje více druhů anomálie, může být zařazen do více kategorií. Kategorie 1 obsahuje lexémy s formální anomálií. Velká část z této skupiny (126 lexémů ze 158) jsou lexémy obsahující neproduktivní afix a/nebo vázanou lexikální bázi. Tuto skupinu tvoří téměř výlučně výpůjčky latinského a řeckého původu (např. *scripture, scientific*) nebo lexémy tvořené v angličtině z přejatých morfémů na základě analogie s těmito výpůjčkami (např. *atonement, dilatory*). Kromě této třídy obsahuje vzorek také 22 lexémů s jinou formální anomálií (např. *elderly, olden, father-in-law*). Kategorie 2 obsahuje lexémy s kolokační anomálií (81 lexémů). Také v této kategorii se vyskytuje podtřída výpůjček z latiny a řečtiny, která se vyznačuje kombinací systematicky používaného afixu (produktivního či neproduktivního) s nesystematicky používanou druhou částí, která se vyskytuje jen ve výpůjčkách. Tato slova jsou tedy na pomezí mezi značkovými a popisnými lexémy. Z 81 lexémů s kolokační anomálií představuje tento podtyp 38 lexémů, tedy přibližně polovinu. Další kolokační anomálie zahrnují jiné monokolokabilní prvky (kromě třídy popsané výše), sémantickou nekompatibilitu a anomální kombinaci z hlediska formálního. Kategorie 3 popisuje nižší stupeň sémantické anomálie, při kterém lexém zachovává svůj pravidelný význam, avšak kromě toho má také význam přenesený (např. *mouth-watering, offshoot, unearth*). Vzorek obsahuje 41 lexémů tohoto druhu. Kategorie 4 zahrnuje lexémy, které mají pouze idiomatický význam, a čítá 128 lexémů. Tato třída se zdá nejvíc specifická formálně, protože na rozdíl od ostatních skupin obsahuje výrazně jiné rozložení slovních druhů i slovotvorných procesů: zastoupení substantiv je výrazně vyšší (86 lexémů, tj. 67,5 % z lexémů zařazených do příslušné kategorie) a také zastoupení kompozit je výrazně vyšší (61 lexémů, tj. 47,6 % z lexémů zařazených do příslušné kategorie).

Výsledky analýzy BNC vzorku vedly k upravení a upřesnění definice lexikálního idiomu. Z analýzy vyplývá, že pokud založíme definici lexikálního idiomu v angličtině na kritériu produktivity v současném jazyce, velká část lexikálních idiomů (v případě BNC vzorku více než 50 % všech anomálních lexémů) bude anomální pouze formálně, a nikoliv sémanticky kvůli velkému zastoupení výpůjček a na základě latiny vytvořených slov v angličtině. Navíc z obecných vlastností slova (vždy do značné míry formálně fixovaného celku, na který působí procesy institucionalizace a lexikalizace) vyplývá, že formální anomálie nad rámec této obecně platné formální stálosti bude méně výrazně vnímaná než podobná formální anomálie na úrovni kombinace slov. Zdá se tedy vhodné přidat na váze kritériu sémantickému. Upravená definice tedy popisuje lexikální idiom jako jednoslovný polymorfémní lexém vyznačující se sémantickou anomálií, která může být doplněna také anomálií formální nebo kolokační. Dále je přesněji definováno, že anomálii určíme na základě analogie s jinými formacemi, a ne na základě produktivity v současném jazyce.

Upravená definice byla použita pro získání druhého vzorku z Oxford English Dictionary (dále OED vzorek). Tento vzorek zahrnuje 500 lexikálních idiomů, které jsou následně analyzovány podobně jako předchozí vzorek (kapitola 6.2.), avšak s větším důrazem na analýzu sémantických vztahů. OED vzorek je omezen pouze na lexémy s prvním zápisem do OED od roku 1800 do současnosti a vznikem slova v rámci angličtiny (vylučuje tedy výpůjčky). Účelem tohoto omezení je zaměřit se na formálně více průhledné formace a naopak vyhnout se různým z etymologických důvodů formálně nejasným lexémům. Z formálního hlediska jsou lexémy analyzovány z hlediska slovního druhu a slovotvorného procesu. Kromě v literatuře zmiňovaných kompozit a odvozenin je také zavedena kategorie kombinovaných formací. Kombinované formace jsou např. syntetická kompozita typu *care-taker*, která vznikla kombinací kompozice a derivace. Poslední formální kategorie zahrnuje atypické kombinace typu křížení (*motel*) nebo mechanického krácení spojeného s dalším slovotvorným procesem (*surfactant*).

První část analýzy OED vzorku se věnuje sémantické anomálii. Slova jsou nejprve podobně jako v prvním vzorku rozdělena na kategorii méně idiomatických slov, která mají více významů, z nichž některé jsou pravidelné a některé idiomatické, a kategorii lexémů pouze

s idiomatickým významem. Vzorek obsahuje 161 lexémů prvního typu, mezi nimiž je poměrně velké zastoupení substantiv i adjektiv (59,6 % substantiv a 31,7 % adjektiv) a zastoupení kompozit a odvozených slov je přibližně vyrovnané (48,4 % kompozit a 42,8 % odvozenin). Druhý typ je zastoupen 339 lexémy a formální složení je poněkud jiné: substantiv je výrazně víc než adjektiv (80,2 % substantiv a 13,6 % adjektiv) a kompozit je výrazně víc než odvozenin (63,7 % kompozit a 24,4 % odvozenin).

Dále jsou prozkoumány základní podtypy sémantické anomálie se zaměřením na neshodu mezi slovotvorným a lexikálním významem. Prvním analyzovaným sémantickým podtypem je specializace významu. Do této třídy jsou zařazeny lexémy, jejichž slovotvorný význam je širší než význam lexikální a zároveň je význam lexikální zahrnut ve významu slovotvorném (např. *output, mobilization, activist*). Specializace významu se vyskytuje u 24,6 % OED vzorku. Tento podtyp je tedy poměrně častý, ale zároveň je velmi problematický z hlediska svého vztahu k idiomatičnosti. Problém spočívá v tom, že sémantická specializace je do určité míry běžnou součástí jazykového pojmenování vznikající jak v momentě tvoření nového slova, tak postupnou specializací při vzniku nového významu existujícího slova. Z tohoto hlediska je tedy sémantická specializace jen okrajovým znakem idiomatičnosti a záleží na stupni a především průhlednosti konkrétního případu specializace. Druhý a třetí typ z hlediska zastoupení jsou typy sémantické neshody často v souvislosti s idiomatickými formacemi zmiňované. Prvním z nich je metafora. Metafora je základem neshody mezi slovotvorným a lexikálním významem u 21 % OED vzorku. Metafora se může vyskytnout v odvozeninách (např. *multidimensional, telling*), v kompozitech (např. *viewpoint, blueprint*) i v kombinovaných formacích (*např. outgrowth, overridding*). Druhým podtypem zmiňovaným častěji v literatuře v souvislosti s idiomatičností jsou exocentrické lexémy. Ve vzorku se jich nachází 10 %. Na rozdíl od tradičního pojetí exocentricity v literatuře jako znaku spojeného s kompozity tato studie zahrnuje do kategorie exocentrických lexémů také odvozeniny, jejichž vztah k referentu je analogický k exocentrickým kompozitům. Příklady exocentrických kompozit z OED vzorku zahrnují *know-nothing, paperback, printout* a odvozeniny jsou reprezentovány např. substantivy *detective, documentary, micro-wave*. Zdá se, že metafora a exocentricita jsou velmi významné podtypy z hlediska idiomatičnosti, protože kromě přímé neshody mezi slovotvorným a lexikálním významem dochází také ke změně domény

v případě metafory a ke změně formální signalizace referenta v případě exocentrických formací. Dalším, již méně zastoupeným podtypem je metonymie (7 %), která je reprezentována např. lexémy *chairperson* a *white-collar*. Zdá se, že metonymie vede k menšímu stupni idiomatičnosti než metafora a excentricita, protože nemění doménu a je v jazyce velmi běžná na více úrovních. Ve vzorku se vyskytují i nepříliš frekventované výskyty jiných sémantických změn, např. zlepšení významu (*exceptional, standout*), zhoršení významu (*collaborator, opportunist*), generalizace významu (*suitcase, poster*), determinizace (*exponentially, fluorescent*).

Kromě výše zmíněných podtypů velká část vzorku (33 %) není zařazena do žádné třídy. Jedná se o lexémy, u nichž je neshoda mezi lexikálním a slovotvorným významem založená na nějakém idiosynkratickém vztahu (např. *hallway, runway, screenplay, black box*).

Následující část studie se věnuje formálním a kolokačním anomáliím. Celkem pouze 22% z OED vzorku obsahuje kromě sémantické anomálie ještě formální a/nebo kolokační anomálii. V práci jsou popsány 4 typy formální anomálie: anomální tvar komponentu (např. *contraceptive, spokesperson*), anomální sufix (např. substantivum *wireless,* sloveso *pressure*), frázová struktura kompozita (*face-to-face*), syntaktická či morfologická anomálie (např. substantivum *vallue-added,* adverbium *overseas*). Z čistě formálních anomálií je nejčastější první typ (4,6 %), hlavně proto, že zahrnuje všechny nepravidelné slovotvorné komponenty, které participují v křížení, mechanickém krácení apod. Ostatní typy se vyskytují velmi okrajově. Na rozhraní mezi formální a kolokační anomálií je formálně anomální kombinace komponentů, která zahrnuje případy kombinací morfémů ze slovotvorného hlediska nepravidelné. Jedná se například o nezvyklé kombinace slovního druhu báze a afixu, který se obvykle pojí s bázemi jiného slovního druhu (např, *steamer, capacitor, insider*), nebo se jedná o nezvyklou kombinaci afixů (*knowledgeable, tailored*). Do určité míry jsou anomální i některé méně prototypické slovotvorné struktury, např. kompozita románského typu (*ceasefire*). OED vzorek obsahuje 5 % takových formací. Typem sémanticky-kolokačním je anomálie spočívající v sémantické nekompatibilitě. Sémantická nekompatibilita vzniká v důsledku sémantické anomálie jednoho nebo obou komponentů, kdy vzniká spojení komponentů, které v základním významu nejsou kompatibilní (*deadline, fast food, airport,*

soap opera). Tento typ se vyskytuje v 6,6 % OED vzorku. Posledním typem, částečně formálním a částečně sémantickým, je tautologie. Tautologie převážně formální se objevuje v případech, kdy se význam báze prakticky shoduje s významem celé kombinace (*dosage, packaging*). Dalším typem je nepravidelné kombinování přípon při pojmenovávání převážně terminologickém (*capacitance, fracionation*). Předposledním typem je sémantická tautologie vyskytující se v kompozitech (*shot-gun, driveway, age-old*). V těchto případech dochází v lexikálním významu ke specializaci, takže výsledný význam není totožný s významem řídícího členu. Tautologie se vyskytuje v 2,8 % lexémů ve vzorku. Posledním typem formální anomálie je nízká až jedinečná kolokabilita vyskytující se v 1 % vzorku (*buffer, sewage*).

Poslední část druhé analýzy se stručně věnuje pragmatickým funkcím lexikálních idiomů zmiňovaným v souvislosti s kolokačními idiomy, tedy expresívní, symbolické a evaluativní složce významu. Na základě vzorku se zdá, že symbolická složka významu je pro některé typy lexikálních idiomů typická a evaluativní složka je poměrně častá (avšak chybí kvantitativní srovnání s kolokačními frazémy). Naproti tomu expresívní složka významu, která je typická pro kolokační frazémy, není u lexikálních nijak výrazně zastoupena.

Nejdůležitější poznatky plynoucí z výše popsané analýzy jsou tyto:

Z formálního hlediska lexikální idiomy nezahrnují pouze třídu kompozit a odvozenin, ale i třídu kombinovaných formací a atypických formací založených na mechanickém krácení forem. Poslední kategorie (a pravděpodobně i předposlední) je typická pro angličtinu a v češtině je mnohem více okrajová. Mechanické krácení je však z hlediska frazeologie také problematické, protože zvláště v ad hoc užití je transparentnost nové formy nezbytná a nedá se tedy mluvit o neprůhledném významu.

Typické slovní druhy lexikálních idiomů jsou substantiva a adjektiva (adjektiva hlavně u typu smíšeného s idiomatickým i neidiomatickým významem). Ze slovotvorných procesů je velmi častá kompozice, i když ani odvozeniny nejsou marginální. Velmi výrazná (108 lexémů) je třída substantiv tvořených kompozicí z frázových sloves (*putoff, printout*).

Naproti tomu některé slovní třídy jsou v kategorii velmi okrajové. Týká se to především adverbií, která se tvoří v naprosté většině příponou *-ly*. Tato přípona je velmi produktivní a

adverbia vytvořená pomocí *-ly* jsou velmi dobře transparentní. Další poměrně málo idiomatické přípony jsou adjektivní *-ing* a *-ed.* Obecněji lze podle dat z obou vzorků říct, že odvozeniny mají větší tendenci k průhlednosti významu, zatímco kompozita mají větší tendenci k neprůhlednosti. Zajímavá je i skupina lexémů vzniklá tzv. zpětným tvořením (*babysit, escalate*). Přestože je zpětné tvoření ze slovotvorného hlediska chápáno jako atypické, z hlediska frazeologie se jedná o pravidelné kombinace, jelikož forma vzniklá zpětným tvořením vzniká právě na základě synchronní analogie.

Na základě provedených analýz práce shrnuje, že existují minimálně tři osy, které určují stupeň idiomatičnosti lexikálního idiomu. První osa směřuje od sémanticky průhledných lexémů přes lexémy v některých významech průhledné a v jiných idiomatické až k lexémům pouze idiomatickým. Druhá osa směřuje od plné kompozicionality přes méně idiomatické podtypy sémantické anomálie až k více idiomatickým podtypům. Třetí osa směřuje od formálně a kolokačně pravidelných lexémů, přes drobnější formální a kolokační anomálie až k nevýraznějším anomáliím typu monokolokability. Následující tabulka ilustruje tuto stupňovitost:

osy idiomatičnosti		**stupně idiomatických vlastností**		
kompozicionalita	průhlednost / neprůhlednost	průhledný	průhledný / neprůhledný	pouze neprůhledný
	neshoda mezi lexikálním a slovotvorným významem	žádná neshoda	lehká neshoda v rámci systematického užití (např. metonymie) až méně systematické, kreativní změny významu (např. metafora)	velké idiosynkratické neshody mezi lexikálním a slovotvorným významem
kombinační a formální pravidelnost / nepravidelnost		formální a kolokační pravidelnost	formální a/nebo kolokační nepravidelnost	velká kombinační anomálie (monokolokabilita)

Na základě dat z obou vzorků se zdá, že jednoslovné lexémy lze skutečně zkoumat v rámci frazeologie, a to především díky hlavní podobnosti s idiomy vyššího stupně: stejně jako

kombinace slov, i polymorfémní lexémy jsou obvykle tvořeny podle určitých pravidel a tato pravidla mohou být v některých případech narušena. Existují však také výrazné rozdíly mezi kombinacemi morfémů a slov: především jsou pravidla pro kombinaci slov mnohem obecněji platná a případné anomálie jsou tedy výraznější, zatímco kombinace morfémů jsou vždy výrazněji omezeny institucionalizací a lexikalizací. Současný výzkum navíc naznačuje, že i pravidelné lexémy jsou v mozku ukládány primárně jako celek a k jejich dekompozici dochází jen v určitých případech. Na základě výše zmíněných argumentů se tedy zdá, že lze třídu lexikálních idiomů zahrnout do studia frazeologie, avšak zároveň je vhodné ji chápat jako oblast periferní s některými vlastními pravidly odlišujícími lexikální idiomy od idiomů vyššího řádu.

Appendix

The first part of the appendix lists all lexemes included in the BNC sample. It contains 1000 lexemes with corresponding data about frequency in the BNC. There are four word-classes: nouns (N), adjectives (ADJ), adverbs (ADV) and verbs (V). Simple lexemes are listed first (SIM) and complex lexemes (COMP) follow. All complex lexemes are then provided with data about the word-formation process involved (C – compounds, D – derivatives, C+D – combined formations), and the type of anomaly (FA 1 – category 1, i.e. formal anomaly, CA 2 – category 2, i.e. collocational anomaly, SA 3 – category 3, i.e. semantic anomaly of lexemes with both regular and idiomatic senses, SA 4 – category 4, i.e. semantic anomaly of lexemes with idiomatic meaning only).

The second part of the appendix contains all lexemes of the OED sample. There are 500 lexical idioms in the second sample. The lexemes are this time sorted according to the word-formation process involved (C – compounds, C+D – combined formations, D – derivatives, other – other formations). The table then includes information about the word-class of the lexeme (it contains the same four open word-classes as the BNC sample), semantic anomaly: SA 4 - semantic anomaly of lexemes with idiomatic meaning only, SA 3 - semantic anomaly of lexemes with both regular and idiomatic senses, FA – formal anomaly, CA – collocational anomaly). Columns SA 4 and SA 3 contain a cross to mark the presence of the anomaly. Subclasses of the formal anomaly are marked in the column by an abbreviation: acf – anomalous combination of forms, as – anomalous syntactic/morphological behaviour, ph – phrasal compounds, ac – anomalous component. Instances of tautology are marked with *t* in the CA column and other collocational anomalies are marked with a cross. Semantic subtype is marked in the next column (metaphor, metonymy, specialization, generalization, exocentric formations, etc.). Lexemes on the borderline between idioms and terms are marked with T in the column named *field-specific*. Particle compounds are marked with P in the next column and evaluative idioms are marked with E in the last column.

Part I – The BNC sample

nr	lexeme	WC	WF	frequency	structure	FA 1	CA 2	SA 3	SA 4
1	privy	ADJ		495	SIM				
2	maxi	ADJ		33	SIM				
3	white	ADJ		19760	SIM				
4	alert	ADJ		699	SIM				
5	burgundy	ADJ		46	SIM				
6	instant	ADJ		1150	SIM				
7	drab	ADJ		280	SIM				
8	weird	ADJ		1054	SIM				
9	bogus	ADJ		327	SIM				
10	whig	ADJ		247	SIM				
11	fresh	ADJ		6745	SIM				
12	lithe	ADJ		111	SIM				
13	rare	ADJ		4876	SIM				
14	how	ADV		98967	SIM				
15	clear	ADV		408	SIM				
16	grimoire	N		34	SIM				
17	figment	N		81	SIM				
18	anthem	N		400	SIM				
19	barrel	N		1406	SIM				
20	catch	N		1039	SIM				
21	viper	N		102	SIM				
22	pop	N		2297	SIM				
23	canton	N		228	SIM				
24	phosphor	N		53	SIM				
25	money	N		36671	SIM				
26	gray	N		1030	SIM				
27	snuff	N		101	SIM				
28	lapis	N		80	SIM				
29	schema	N		460	SIM				
30	autobahn	N		70	SIM				
31	jacquard	N		177	SIM				
32	angora	N		40	SIM				
33	kitchen	N		8211	SIM				
34	cashew	N		31	SIM				
35	muderris	N		47	SIM				
36	nit	N		68	SIM				
37	peace	N		8660	SIM				
38	plexus	N		72	SIM				
39	pasty	N		65	SIM				
40	doyen	N		63	SIM				
41	parable	N		374	SIM				
42	aspirin	N		354	SIM				
43	sheikh	N		348	SIM				
44	time	N		180243	SIM				
45	rugby	N		3433	SIM				
46	catarrh	N		67	SIM				
47	penis	N		514	SIM				
48	pheasant	N		317	SIM				
49	threat	N		6903	SIM				
50	spit	N		262	SIM				
51	ganglion	N		220	SIM				
52	jacuzzi	N		78	SIM				
53	diktat	N		30	SIM				

54	pass	N		2347	SIM				
55	entablature	N		32	SIM				
56	croat	N		474	SIM				
57	nib	N		72	SIM				
58	haw	N		30	SIM				
59	twist	N		935	SIM				
60	alcove	N		178	SIM				
61	balustrade	N		121	SIM				
62	crouch	N		118	SIM				
63	kiosk	N		236	SIM				
64	chandelier	N		200	SIM				
65	shark	N		543	SIM				
66	pleb	N		41	SIM				
67	fiesta	N		193	SIM				
68	trill	N		63	SIM				
69	closet	N		235	SIM				
70	disc	N		2352	SIM				
71	arc	N		976	SIM				
72	fringe	N		1219	SIM				
73	nucleus	N		965	SIM				
74	sort	N		28003	SIM				
75	vagabond	N		81	SIM				
76	roulette	N		62	SIM				
77	ferry	N		1447	SIM				
78	story	N		17791	SIM				
79	pinion	N		31	SIM				
80	dawn	N		2237	SIM				
81	git	N		235	SIM				
82	paranoia	N		208	SIM				
83	soil	N		4723	SIM				
84	mufti	N		255	SIM				
85	weasel	N		157	SIM				
86	vest	N		360	SIM				
87	gilt	N		610	SIM				
88	egg	N		6064	SIM				
89	blossom	N		416	SIM				
90	crease	N		204	SIM				
91	apostle	N		448	SIM				
92	cade	N		51	SIM				
93	marvel	N		162	SIM				
94	bladder	N		1046	SIM				
95	prison	N		7049	SIM				
96	splint	N		67	SIM				
97	brandy	N		882	SIM				
98	galleon	N		78	SIM				
99	noise	N		5280	SIM				
100	maze	N		491	SIM				
101	haemorrhoid	N		30	SIM				
102	rabies	N		80	SIM				
103	jig	N		167	SIM				
104	tabard	N		40	SIM				
105	trench	N		895	SIM				
106	tassel	N		87	SIM				
107	chutney	N		67	SIM				
108	glance	N		2429	SIM				
109	diaphragm	N		157	SIM				
110	groin	N		354	SIM				
111	hope	N		8406	SIM				
112	list	N		13661	SIM				

113	anemone	N		232	SIM				
114	conflict	N		6970	SIM				
115	camomile	N		50	SIM				
116	vesicle	N		139	SIM				
117	ice	N		3981	SIM				
118	quail	N		125	SIM				
119	enigma	N		249	SIM				
120	trestle	N		140	SIM				
121	throne	N		1256	SIM				
122	spleen	N		151	SIM				
123	genie	N		64	SIM				
124	plasma	N		981	SIM				
125	harp	N		300	SIM				
126	lentil	N		113	SIM				
127	gristle	N		31	SIM				
128	toe	N		1610	SIM				
129	suspense	N		200	SIM				
130	dessert	N		437	SIM				
131	shampoo	N		346	SIM				
132	capsule	N		385	SIM				
133	lint	N		37	SIM				
134	reproach	N		165	SIM				
135	quorum	N		147	SIM				
136	banister	N		140	SIM				
137	chaplain	N		539	SIM				
138	mode	N		3929	SIM				
139	audio	N		188	SIM				
140	orgasm	N		215	SIM				
141	emphysema	N		46	SIM				
142	cog	N		110	SIM				
143	slurry	N		180	SIM				
144	lynx	N		140	SIM				
145	digest	N		299	SIM				
146	ghetto	N		281	SIM				
147	creak	N		109	SIM				
148	republic	N		5694	SIM				
149	angel	N		2240	SIM				
150	fix	N		283	SIM				
151	hullabaloo	N		36	SIM				
152	academy	N		1448	SIM				
153	lead	N		5728	SIM				
154	boomerang	N		35	SIM				
155	orchestra	N		1677	SIM				
156	grace	N		2421	SIM				
157	bulletin	N		984	SIM				
158	bough	N		183	SIM				
159	flab	N		32	SIM				
160	delicatessen	N		134	SIM				
161	odour	N		924	SIM				
162	duct	N		404	SIM				
163	rota	N		194	SIM				
164	measure	N		11092	SIM				
165	grin	N		1095	SIM				
166	scorn	N		297	SIM				
167	damsel	N		103	SIM				
168	zeta	N		199	SIM				
169	plaudit	N		52	SIM				
170	repartee	N		39	SIM				
171	bismuth	N		99	SIM				

172	area	N		58029	SIM				
173	deer	N		909	SIM				
174	january	N		10096	SIM				
175	insect	N		2103	SIM				
176	brawn	N		37	SIM				
177	crown	N		5312	SIM				
178	mote	N		52	SIM				
179	finance	N		6234	SIM				
180	stool	N		1087	SIM				
181	move	N		8046	SIM				
182	esprit	N		82	SIM				
183	rain	N		6127	SIM				
184	carbon	N		2486	SIM				
185	doubt	N		11764	SIM				
186	console	N		212	SIM				
187	confidante	N		63	SIM				
188	fruit	N		4985	SIM				
189	cargo	N		1011	SIM				
190	groan	N		318	SIM				
191	canine	N		60	SIM				
192	gloss	N		394	SIM				
193	abstract	N		370	SIM				
194	beaver	N		195	SIM				
195	mayor	N		2377	SIM				
196	limbo	N		178	SIM				
197	nervosa	N		176	SIM				
198	ode	N		149	SIM				
199	chase	N		940	SIM				
200	bustard	N		40	SIM				
201	gutter	N		553	SIM				
202	example	N		43028	SIM				
203	parody	N		286	SIM				
204	mirza	N		41	SIM				
205	contact	N		8553	SIM				
206	stanchion	N		49	SIM				
207	swire	N		68	SIM				
208	sexton	N		91	SIM				
209	prefect	N		111	SIM				
210	journey	N		5380	SIM				
211	neurone	N		192	SIM				
212	rapport	N		295	SIM				
213	serge	N		109	SIM				
214	savoury	N		43	SIM				
215	coal	N		5311	SIM				
216	jab	N		123	SIM				
217	gang	N		1984	SIM				
218	soma	N		30	SIM				
219	draft	N		2773	SIM				
220	carcass	N		251	SIM				
221	quay	N		532	SIM				
222	jabber	V		32	SIM				
223	augment	V		504	SIM				
224	twine	V		72	SIM				
225	ground	V		403	SIM				
226	whine	V		276	SIM				
227	rebut	V		106	SIM				
228	wind	V		2204	SIM				
229	splinter	V		114	SIM				
230	right	V		124	SIM				

231	queue	V		548	SIM				
232	poke	V		681	SIM				
233	prowl	V		182	SIM				
234	stencil	V		45	SIM				
235	mass	V		134	SIM				
236	vest	V		485	SIM				
237	relieve	V		1411	SIM				
238	repeal	V		325	SIM				
239	bankrupt	V		82	SIM				
240	grit	V		290	SIM				
241	absent	V		120	SIM				
242	archive	V		174	SIM				
243	collide	V		561	SIM				
244	prepare	V		10837	SIM				
245	crouch	V		805	SIM				
246	abandon	V		4332	SIM				
247	cushion	V		191	SIM				
248	occur	V		15430	SIM				
249	heckle	V		56	SIM				
250	shall	V		20011	SIM				
251	combine	V		5868	SIM				
252	crunch	V		256	SIM				
253	plate	V		90	SIM				
254	inch	V		284	SIM				
255	got	V		48	SIM				
256	pronounce	V		1132	SIM				
257	impart	V		371	SIM				
258	sip	V		871	SIM				
259	lick	V		883	SIM				
260	idle	V		122	SIM				
261	hall	V		90	SIM				
262	impact	V		93	SIM				
263	annoy	V		579	SIM				
264	cream	V		79	SIM				
265	dab	V		214	SIM				
266	size	V		233	SIM				
267	select	V		5730	SIM				
268	ward	V		189	SIM				
269	lever	V		141	SIM				
270	sederunt	V		30	SIM				
271	fudge	V		76	SIM				
272	scoff	V		182	SIM				
273	defend	V		4145	SIM				
274	ship	V		1589	SIM				
275	quack	V		60	SIM				
276	cleave	V		132	SIM				
277	destine	V		767	SIM				
278	accept	V		19811	SIM				
279	conquer	V		599	SIM				
280	brood	V		203	SIM				
281	employ	V		7826	SIM				
282	indulge	V		985	SIM				
283	founder	V		227	SIM				
284	brief	V		628	SIM				
285	nerve	V		51	SIM				
286	crackle	V		220	SIM				
287	revere	V		33	SIM				
288	dip	V		1146	SIM				
289	hinder	V		594	SIM				

290	shunt	V		156	SIM				
291	bunk	V		39	SIM				
292	numb	V		110	SIM				
293	anger	V		877	SIM				
294	wan	V		2616	SIM				
295	knife	V		62	SIM				
296	vary	V		6267	SIM				
297	ransom	V		64	SIM				
298	picnic	V		42	SIM				
299	think	V		145438	SIM				
300	spice	V		114	SIM				
301	ice	V		122	SIM				
302	hand	V		5095	SIM				
303	shave	V		636	SIM				
304	amaze	V		295	SIM				
305	intend	V		10571	SIM				
306	perplex	V		121	SIM				
307	borrow	V		3000	SIM				
308	shock	V		821	SIM				
309	stilt	V		46	SIM				
310	bowl	V		710	SIM				
311	batten	V		51	SIM				
312	cast	V		3569	SIM				
313	pull	V		13103	SIM				
314	reap	V		489	SIM				
315	dub	V		546	SIM				
316	stack	V		662	SIM				
317	swivel	V		287	SIM				
318	stage	V		1585	SIM				
319	brook	V		73	SIM				
320	low-voltage	ADJ	C	42	COMP				
321	sawn-off	ADJ	C	57	COMP			X	
322	enchanted	ADJ	D	140	COMP				
323	garrulous	ADJ	D	40	COMP	X	X		
324	binding	ADJ	D	1222	COMP				
325	superficial	ADJ	D	758	COMP	X	X		
326	horrified	ADJ	D	532	COMP				
327	cross-examined	ADJ	C+D	37	COMP				X
328	columnar	ADJ	D	93	COMP				
329	astral	ADJ	D	42	COMP	X			
330	unrecognizable	ADJ	D	52	COMP				
331	diagrammatic	ADJ	D	81	COMP	X			
332	dose-response	ADJ	C	32	COMP				
333	obliged	ADJ	D	1631	COMP			X	
334	archaeological	ADJ	D	874	COMP				
335	illustrated	ADJ	D	301	COMP				
336	weightless	ADJ	D	59	COMP			X	
337	audiovisual	ADJ	C+D	58	COMP				
338	creaking	ADJ	D	164	COMP				
339	updated	ADJ	D	252	COMP				
340	mouth-watering	ADJ	C+D	45	COMP			X	
341	literate	ADJ	D	341	COMP	X	X		
342	ready-made	ADJ	C	246	COMP			X	
343	stroppy	ADJ	D	37	COMP				
344	silent	ADJ	D	3489	COMP	X			
345	dogged	ADJ	D	109	COMP		X		X
346	manometric	ADJ	C+D	72	COMP				
347	severed	ADJ	D	129	COMP				
348	cleft	ADJ	D	43	COMP	X			

349	self-assured	ADJ	C+D	45	COMP				
350	nippy	ADJ	D	44	COMP				
351	matrilineal	ADJ	C+D	51	COMP				
352	atypical	ADJ	D	144	COMP	X			
353	gilt-edged	ADJ	C+D	96	COMP			X	
354	landward	ADJ	D	42	COMP				
355	exacting	ADJ	D	178	COMP				
356	incongruous	ADJ	D	167	COMP	X			
357	reasoned	ADJ	D	198	COMP				X
358	aware	ADJ	D	10464	COMP	X			
359	defunct	ADJ	D	154	COMP	X	X		X
360	removable	ADJ	D	135	COMP				
361	shrivelled	ADJ	D	42	COMP				
362	pulling	ADJ	D	37	COMP				
363	tubby	ADJ	D	43	COMP			X	
364	alarmed	ADJ	D	436	COMP				
365	national	ADJ	D	37463	COMP				
366	extra-parliamentary	ADJ	D	37	COMP				
367	disorderly	ADJ	D	202	COMP	X			
368	unemotional	ADJ	D	55	COMP				
369	antenatal	ADJ	D	152	COMP				
370	record-breaking	ADJ	C+D	97	COMP				
371	purposeful	ADJ	D	264	COMP	X			
372	cautious	ADJ	D	1099	COMP	X			
373	anti-jewish	ADJ	D	30	COMP				
374	unwise	ADJ	D	401	COMP				
375	top-level	ADJ	C	77	COMP				X
376	high-density	ADJ	C+D	34	COMP				
377	unrealised	ADJ	D	54	COMP				
378	greyish	ADJ	D	64	COMP				
379	cynical	ADJ	D	740	COMP				
380	respective	ADJ	D	1215	COMP	X			X
381	unrealistic	ADJ	D	577	COMP				
382	inductive	ADJ	D	104	COMP	X			
383	illicit	ADJ	D	261	COMP				
384	episcopal	ADJ	D	251	COMP	X	X		
385	well-designed	ADJ	C+D	77	COMP				
386	untidy	ADJ	D	382	COMP				
387	attacking	ADJ	D	210	COMP				
388	extra-mural	ADJ	D	125	COMP				X
389	five-door	ADJ	C	34	COMP				
390	cylindrical	ADJ	D	182	COMP				
391	teleological	ADJ	D	53	COMP				
392	histopathological	ADJ	D	35	COMP				
393	sclerosing	ADJ	D	74	COMP				
394	linoleic	ADJ	D	31	COMP				
395	hypoglycaemic	ADJ	D	41	COMP				
396	mystic	ADJ	D	120	COMP	X			
397	free-swimming	ADJ	C+D	41	COMP				
398	fractious	ADJ	D	58	COMP	X			X
399	shamefaced	ADJ	C+D	51	COMP				X
400	19-year-old	ADJ	C	142	COMP				
401	revealing	ADJ	D	381	COMP				
402	skinny	ADJ	D	315	COMP			X	
403	preserving	ADJ	D	146	COMP				
404	islamic	ADJ	D	1294	COMP				
405	long-eared	ADJ	C+D	49	COMP				
406	caught	ADJ	D	45	COMP				
407	moronic	ADJ	D	32	COMP				

408	non-stop	ADJ	D	143	COMP				
409	obligatory	ADJ	D	325	COMP	X			
410	temperate	ADJ	D	404	COMP	X	X		
411	curly	ADJ	D	411	COMP				
412	tanning	ADJ	D	32	COMP				
413	fallacious	ADJ	D	59	COMP	X			
414	allegorical	ADJ	D	92	COMP				
415	bonded	ADJ	D	89	COMP				
416	reserve	ADJ	D	826	COMP	X	X		
417	sleepless	ADJ	D	189	COMP				
418	permeable	ADJ	D	68	COMP		1		
419	therapeutic	ADJ	D	673	COMP	X			
420	newfound	ADJ	C	35	COMP	X	X		
421	hypertrophic	ADJ	D	55	COMP				
422	foolproof	ADJ	C	58	COMP				X
423	underdeveloped	ADJ	D	111	COMP				
424	pluralist	ADJ	D	353	COMP				X
425	submissive	ADJ	D	138	COMP	X			
426	radiological	ADJ	D	170	COMP				
427	amazonian	ADJ	D	109	COMP				
428	herculean	ADJ	D	35	COMP				
429	constituent	ADJ	D	57	COMP	X			
430	questioning	ADJ	D	164	COMP			X	
431	disobedient	ADJ	D	72	COMP				
432	exhaustive	ADJ	D	328	COMP				
433	herbal	ADJ	D	199	COMP				
434	money-making	ADJ	C+D	36	COMP				
435	retiring	ADJ	D	267	COMP			X	
436	risky	ADJ	D	705	COMP				
437	zimbabwean	ADJ	D	92	COMP				
438	government-owned	ADJ	C+D	39	COMP				
439	mental	ADJ	D	5714	COMP	X			
440	smashed	ADJ	D	145	COMP			X	
441	well-documented	ADJ	C+D	85	COMP				
442	asymmetric	ADJ	D	135	COMP	X			
443	sicilian	ADJ	D	134	COMP				
444	lacquered	ADJ	D	45	COMP				
445	traditional	ADJ	D	9696	COMP				
446	institutionalized	ADJ	D	88	COMP				X
447	banded	ADJ	D	58	COMP				
448	dual	ADJ	D	1140	COMP	X			
449	wholesale	ADJ	C	682	COMP		X		X
450	topmost	ADJ	C	73	COMP	X			
451	duplicate	ADJ	D	120	COMP	X	X		
452	antisocial	ADJ	D	50	COMP				
453	wimpy	ADJ	D	32	COMP				
454	joyless	ADJ	D	45	COMP				
455	decreasing	ADJ	D	228	COMP				
456	usual	ADJ	D	7328	COMP	X			X
457	thorny	ADJ	D	143	COMP			X	
458	supercilious	ADJ	D	50	COMP	X	X		
459	braided	ADJ	D	63	COMP				
460	flexible	ADJ	D	2379	COMP				X
461	unprotected	ADJ	D	175	COMP				
462	slippy	ADJ	D	34	COMP				
463	self-indulgent	ADJ	C+D	101	COMP				
464	polarised	ADJ	D	85	COMP				
465	bad-tempered	ADJ	C+D	100	COMP				
466	nervous	ADJ	D	2889	COMP				

467	devastated	ADJ	D	57	COMP				
468	filmed	ADJ	D	37	COMP				
469	elderly	ADJ	D	4899	COMP	X			
470	differentiated	ADJ	D	94	COMP				
471	required	ADJ	D	1975	COMP				
472	offensive	ADJ	D	931	COMP	X			
473	on-board	ADJ	C	95	COMP				X
474	unflattering	ADJ	D	72	COMP				
475	venerable	ADJ	D	182	COMP	X			
476	needed	ADJ	D	34	COMP				
477	godly	ADJ	D	98	COMP	X			X
478	isotopic	ADJ	C	93	COMP				
479	desirable	ADJ	D	2077	COMP				
480	understanding	ADJ	D	49	COMP				
481	multi-level	ADJ	D	40	COMP				
482	brilliant	ADJ	D	3398	COMP	X			
483	computer-aided	ADJ	C+D	148	COMP				
484	uncorrelated	ADJ	D	45	COMP				
485	paralytic	ADJ	D	40	COMP	X			
486	coldwater	ADJ	C	64	COMP				X
487	far-away	ADJ	C	36	COMP	X			
488	unreasonable	ADJ	D	978	COMP				
489	serial	ADJ	D	636	COMP	X			
490	arcaded	ADJ	D	42	COMP				
491	unacknowledged	ADJ	D	81	COMP				
492	time-dependent	ADJ	C+D	30	COMP				
493	persuasive	ADJ	D	535	COMP	X			
494	countrywide	ADJ	D	53	COMP				
495	adenomatous	ADJ	D	95	COMP				
496	especial	ADJ	D	128	COMP	X			
497	confusing	ADJ	D	774	COMP				
498	paediatric	ADJ	D	119	COMP				
499	proven	ADJ	D	450	COMP			X	
500	animate	ADJ	D	59	COMP	X	X		
501	coated	ADJ	D	63	COMP				
502	undocumented	ADJ	D	33	COMP				
503	dental	ADJ	D	610	COMP	X			
504	salient	ADJ	D	332	COMP	X	X		
505	dimensional	ADJ	D	189	COMP				
506	self-regulating	ADJ	D	75	COMP				
507	encircling	ADJ	D	38	COMP				
508	spiked	ADJ	D	35	COMP				
509	symmetrical	ADJ	D	329	COMP				
510	burned	ADJ	D	67	COMP				
511	despotic	ADJ	D	39	COMP				
512	prepatent	ADJ	D	52	COMP				X
513	non-linear	ADJ	D	104	COMP				
514	blackened	ADJ	D	170	COMP				
515	contrary	ADJ	D	1626	COMP	X			
516	insecure	ADJ	D	329	COMP				
517	purported	ADJ	D	90	COMP				
518	two-part	ADJ	C	107	COMP				
519	black-headed	ADJ	C+D	31	COMP				
520	coniferous	ADJ	D	68	COMP				
521	contemplative	ADJ	D	216	COMP	X			
522	warring	ADJ	D	188	COMP				
523	scientific	ADJ	D	5796	COMP	X			
524	scratchy	ADJ	D	55	COMP				
525	unspecific	ADJ	D	33	COMP				

526	big-name	ADJ	C	36	COMP				
527	unacceptable	ADJ	D	1219	COMP				
528	impartial	ADJ	D	309	COMP				
529	bestial	ADJ	D	49	COMP	X			
530	timbered	ADJ	D	35	COMP				
531	sensible	ADJ	D	2677	COMP				X
532	coastal	ADJ	D	1423	COMP				
533	unrelated	ADJ	D	561	COMP				
534	self-deprecating	ADJ	C+D	37	COMP				
535	polyunsaturated	ADJ	D	47	COMP		X		
536	attributable	ADJ	D	582	COMP				
537	practising	ADJ	D	486	COMP				
538	discredited	ADJ	D	126	COMP				
539	unsympathetic	ADJ	D	155	COMP				
540	all-night	ADJ	C	120	COMP				
541	uneconomical	ADJ	D	51	COMP				
542	fitted	ADJ	D	298	COMP				
543	ill-tempered	ADJ	C+D	33	COMP				
544	olden	ADJ	D	52	COMP	X			
545	binomial	ADJ	D	47	COMP	X			
546	mindful	ADJ	D	174	COMP				X
547	humming	ADJ	D	33	COMP				
548	waste-paper	ADJ	C	34	COMP				
549	synchronic	ADJ	D	30	COMP				
550	selectable	ADJ	D	33	COMP				
551	targeted	ADJ	D	83	COMP				
552	unprofitable	ADJ	D	162	COMP				
553	glassy	ADJ	D	184	COMP			X	
554	under-represented	ADJ	D	88	COMP				
555	unicameral	ADJ	D	121	COMP				
556	underhand	ADJ	C	57	COMP	X			X
557	pre-exposed	ADJ	D	30	COMP				
558	unconstrained	ADJ	D	73	COMP				
559	unattainable	ADJ	D	102	COMP				
560	air-conditioning	ADJ	C+D	84	COMP				
561	well-lit	ADJ	C	49	COMP				
562	fired	ADJ	D	40	COMP				
563	infuriated	ADJ	D	42	COMP				
564	double-sided	ADJ	C+D	60	COMP				X
565	suited	ADJ	D	89	COMP			X	
566	catching	ADJ	D	39	COMP				X
567	gurgling	ADJ	D	50	COMP				
568	autumnal	ADJ	D	71	COMP				
569	inside	ADJ	C	577	COMP				X
570	nonsensical	ADJ	D	90	COMP				
571	pre-christmas	ADJ	D	72	COMP				
572	ornithological	ADJ	D	58	COMP				
573	graphical	ADJ	D	645	COMP				
574	farming	ADJ	D	39	COMP				
575	transitive	ADJ	D	53	COMP	X			X
576	deep-water	ADJ	C	36	COMP				
577	ante-natal	ADJ	D	53	COMP				
578	stilted	ADJ	D	58	COMP			X	
579	whipping	ADJ	D	33	COMP				
580	dilatory	ADJ	D	37	COMP	X			
581	searing	ADJ	D	148	COMP				
582	devoted	ADJ	D	1202	COMP			X	
583	good-natured	ADJ	C+D	125	COMP				
584	swiftly	ADV	D	1169	COMP				

585	precisely	ADV	D	3423	COMP				
586	unequally	ADV	D	40	COMP				
587	crudely	ADV	D	233	COMP				
588	tremendously	ADV	D	298	COMP				
589	permanently	ADV	D	1206	COMP				
590	intensively	ADV	D	179	COMP				
591	smoothly	ADV	D	946	COMP				
592	painlessly	ADV	D	66	COMP				
593	best	ADV	D	7924	COMP	X	X		
594	underway	ADV	C	582	COMP	X			X
595	audibly	ADV	D	46	COMP				
596	fitfully	ADV	D	53	COMP				
597	poetically	ADV	D	33	COMP				
598	commonly	ADV	D	2532	COMP				
599	authentically	ADV	D	60	COMP				
600	vociferously	ADV	D	54	COMP				
601	presumably	ADV	D	3196	COMP				
602	methodologically	ADV	D	32	COMP				
603	honourably	ADV	D	60	COMP				
604	regrettably	ADV	D	228	COMP				
605	dryly	ADV	D	77	COMP				
606	opposite	ADV	D	660	COMP	X			
607	gingerly	ADV	D	209	COMP				X
608	innately	ADV	D	48	COMP				
609	visually	ADV	D	576	COMP				
610	hereby	ADV	C	258	COMP		X		X
611	arguably	ADV	D	627	COMP				
612	diagrammatically	ADV	D	48	COMP				
613	disgustingly	ADV	D	32	COMP				
614	henceforth	ADV	C	329	COMP		X		
615	knowingly	ADV	D	314	COMP				
616	euphemistically	ADV	D	53	COMP				
617	autonomously	ADV	D	41	COMP				
618	testily	ADV	D	63	COMP				
619	laterally	ADV	D	124	COMP				
620	belligerently	ADV	D	37	COMP				
621	irritably	ADV	D	256	COMP				
622	fixedly	ADV	D	65	COMP				
623	neutrally	ADV	D	45	COMP				
624	abroad	ADV	D	3871	COMP				X
625	intellectually	ADV	D	316	COMP				
626	purposefully	ADV	D	142	COMP				
627	grandly	ADV	D	90	COMP				
628	dizzily	ADV	D	30	COMP				
629	openly	ADV	D	1194	COMP				
630	loyally	ADV	D	77	COMP				
631	trading	N	D	3842	COMP				
632	sociolinguist	N	C+D	62	COMP				
633	exhortation	N	D	190	COMP				
634	eurovision	N	C	36	COMP		X		X
635	octagon	N	C	77	COMP				
636	idiocy	N	D	49	COMP	X			
637	oddity	N	D	216	COMP			X	
638	litigation	N	D	835	COMP				
639	foreleg	N	C	54	COMP				
640	shuffling	N	D	41	COMP				
641	crystallisation	N	D	45	COMP				
642	grand-daughter	N	C	77	COMP				X
643	offshoot	N	C	148	COMP			X	

644	scriptwriter	N	C+D	57	COMP				
645	ill-health	N	C	268	COMP				
646	byway	N	C	53	COMP				X
647	insolence	N	D	55	COMP	X			
648	flying-boat	N	C+D	39	COMP				X
649	foreman	N	C	567	COMP				X
650	heckler	N	D	41	COMP				
651	milkman	N	C	209	COMP				
652	aquatint	N	C	54	COMP				
653	patrimony	N	C	100	COMP	X			
654	anarchist	N	D	204	COMP				
655	consumerism	N	D	184	COMP	X	X		X
656	finery	N	D	84	COMP				X
657	leftist	N	D	154	COMP				X
658	set-back	N	C	62	COMP				X
659	protector	N	D	417	COMP			X	
660	republican	N	D	1674	COMP				
661	tonnage	N	D	105	COMP	X			X
662	walling	N	D	110	COMP		X		
663	pounding	N	D	304	COMP				
664	insomniac	N	D	30	COMP	X			
665	humiliation	N	D	616	COMP				
666	logon	N	C	30	COMP				X
667	camellia	N	D	61	COMP				
668	unloading	N	D	33	COMP				
669	hypotension	N	D	39	COMP				
670	roll-out	N	C	40	COMP				X
671	incentive	N	D	2312	COMP	X	X		
672	carving	N	D	477	COMP				
673	dreadnought	N	C	106	COMP				X
674	proprietor	N	D	870	COMP	X	X		
675	traditionalist	N	D	136	COMP				
676	tunnelling	N	D	72	COMP				
677	caring	N	D	403	COMP				
678	typist	N	D	199	COMP				
679	argentinian	N	D	36	COMP				
680	waiting	N	D	1002	COMP				
681	pullover	N	C	163	COMP				X
682	checkpoint	N	C	169	COMP				X
683	abolitionist	N	D	232	COMP			X	
684	reversal	N	D	614	COMP	X			
685	dismemberment	N	D	42	COMP		X		X
686	overthrow	N	C	285	COMP				X
687	lookout	N	C	180	COMP				X
688	galatian	N	D	57	COMP				
689	reprocessing	N	D	260	COMP				X
690	cadence	N	D	92	COMP	X	X		
691	biplane	N	D	85	COMP		X		X
692	supermarket	N	D	1593	COMP				X
693	stimulant	N	D	135	COMP	X			X
694	pusher	N	D	93	COMP			X	
695	registrar	N	D	753	COMP	X	X		
696	claimant	N	D	722	COMP	X			X
697	aerospace	N	C	774	COMP				
698	multi-media	N	D	128	COMP				X
699	cabinet	N	D	6761	COMP				X
700	inception	N	D	289	COMP				X
701	atonement	N	D	134	COMP	X			
702	sociologist	N	D	726	COMP				

703	campaigning	N	D	149	COMP				
704	rhetoric	N	D	955	COMP		X		
705	cleansing	N	D	265	COMP				
706	wrangling	N	D	101	COMP				
707	despotism	N	D	92	COMP				
708	platitude	N	D	80	COMP	X	X		
709	onset	N	C	806	COMP		X		X
710	principal	N	D	1178	COMP	X	X		
711	education	N	D	25873	COMP				
712	re-export	N	D	30	COMP				
713	indicative	N	D	56	COMP				
714	semitone	N	D	69	COMP				
715	carbonate	N	D	601	COMP				
716	honour	N	D	3492	COMP	X	X		
717	bender	N	D	61	COMP			X	
718	welder	N	D	65	COMP				
719	dimer	N	D	70	COMP				
720	historicity	N	D	53	COMP				
721	listeria	N	D	46	COMP				
722	tirade	N	D	103	COMP	X	X		
723	conceptualisation	N	D	40	COMP				
724	scripture	N	D	670	COMP	X			
725	tapping	N	D	95	COMP				
726	deadlock	N	C	277	COMP		X		X
727	defense	N	D	206	COMP	X	X		
728	mistletoe	N	C	112	COMP	X	X		X
729	father-in-law	N	C	202	COMP	X	X		X
730	caerulein	N	D	86	COMP				
731	filename	N	C	265	COMP				
732	tritium	N	D	84	COMP				
733	choreography	N	C	107	COMP				
734	stompie	N	D	33	COMP	X			X
735	sniffer	N	D	66	COMP			X	
736	savagery	N	D	151	COMP				
737	steamboat	N	C	34	COMP				
738	granary	N	D	72	COMP	X			
739	renunciation	N	D	155	COMP	X			
740	south-west	N	C	919	COMP				
741	nihilism	N	D	34	COMP				
742	imagining	N	D	82	COMP				
743	robber	N	D	427	COMP				
744	cross-fertilisation	N	D	34	COMP				
745	bicarbonate	N	D	232	COMP				
746	gunwale	N	C	39	COMP	X			X
747	data-base	N	C	65	COMP		X		X
748	physiology	N	C	330	COMP				
749	heritability	N	D	36	COMP				
750	respondent	N	D	1602	COMP	X			X
751	upkeep	N	C	159	COMP		X		X
752	goer	N	D	47	COMP			X	
753	neocortex	N	D	36	COMP				
754	whirlpool	N	C	146	COMP				X
755	whisker	N	D	254	COMP				X
756	penalty	N	D	3778	COMP	X			
757	punisher	N	D	54	COMP			X	
758	skimmer	N	D	86	COMP				
759	slackening	N	D	31	COMP				
760	analgesic	N	D	97	COMP	X			
761	tipper	N	D	50	COMP	X		X	

762	dysphagia	N	D	76	COMP				
763	tranquillity	N	D	309	COMP				
764	creatinine	N	D	106	COMP				
765	snowball	N	C	137	COMP			X	
766	bromide	N	D	85	COMP				
767	dryness	N	D	123	COMP				
768	side-effect	N	C	304	COMP				X
769	semi-desert	N	D	35	COMP				
770	self-consciousness	N	C+D	145	COMP				
771	rattlesnake	N	C	54	COMP				X
772	innervation	N	D	51	COMP				
773	seascape	N	C	47	COMP		X		X
774	understudy	N	C	99	COMP		X		X
775	additionality	N	D	44	COMP				
776	spendthrift	N	C	33	COMP		X		X
777	goldmine	N	C	46	COMP			X	
778	routing	N	D	80	COMP				
779	payroll	N	C	404	COMP		X		X
780	enjoyment	N	D	1030	COMP	X			
781	commissariat	N	D	50	COMP	X			
782	showbiz	N	C	168	COMP	X	X		X
783	paling	N	D	33	COMP				
784	meeting	N	D	21209	COMP				
785	hypothesis	N	D	2239	COMP				X
786	coldness	N	D	216	COMP				
787	lessor	N	D	181	COMP	X			
788	arabic	N	D	300	COMP				
789	growing	N	D	85	COMP				
790	livestock	N	C	850	COMP		X		X
791	wariness	N	D	63	COMP				
792	solution	N	D	9322	COMP	X	X		
793	patency	N	D	59	COMP	X			X
794	casualty	N	D	1720	COMP	X			X
795	shaper	N	D	47	COMP				
796	appellant	N	D	683	COMP	X			
797	pedestrian	N	D	846	COMP	X	X		
798	incrementalism	N	D	32	COMP				
799	liquidation	N	D	476	COMP				
800	rainbow	N	C	1046	COMP		X		X
801	action	N	D	26481	COMP				
802	histamine	N	D	235	COMP				
803	milling	N	D	201	COMP			X	
804	eagerness	N	D	217	COMP				
805	disinclination	N	D	48	COMP				
806	centreline	N	C	37	COMP				
807	rationalism	N	D	137	COMP				
808	workload	N	C	529	COMP		X		X
809	drive-in	N	C	32	COMP				X
810	disintegration	N	D	318	COMP				
811	peritonitis	N	D	32	COMP				
812	catalan	N	D	65	COMP	X			
813	workbook	N	C	117	COMP			X	
814	haulier	N	D	59	COMP	X			X
815	destructiveness	N	D	38	COMP				
816	bloodstock	N	C	44	COMP		X		X
817	stepping	N	D	153	COMP				
818	stimulation	N	D	811	COMP	X			
819	hacking	N	D	53	COMP				
820	leadership	N	C	4800	COMP				

821	affiliate	N	D	218	COMP	X	X		
822	vicarage	N	D	234	COMP	X			
823	cholecystectomy	N	C	134	COMP				
824	shorthand	N	C	312	COMP				X
825	physiologist	N	C+D	85	COMP				
826	artist	N	D	7799	COMP				
827	paedophile	N	C	31	COMP				X
828	refutation	N	D	88	COMP				
829	disbelief	N	D	598	COMP				
830	fitment	N	D	51	COMP	X			X
831	drawing	N	D	4690	COMP				
832	lexicographer	N	C+D	200	COMP				
833	thatcherism	N	D	225	COMP				
834	mulberry	N	D	122	COMP	X	X		X
835	jackdaw	N	C	73	COMP		X		
836	flagpole	N	C	35	COMP				
837	diversion	N	D	664	COMP	X			X
838	compstation	N	C	30	COMP	X	X		X
839	floorboard	N	C	309	COMP				
840	weighing	N	D	202	COMP				
841	mugger	N	D	142	COMP				
842	screwdriver	N	C+D	269	COMP				
843	collegiality	N	D	52	COMP				X
844	goalkeeping	N	C+D	77	COMP		X		X
845	recapture	N	D	40	COMP				
846	dumping	N	D	276	COMP				
847	flipper	N	D	60	COMP				X
848	goalmouth	N	C	51	COMP		X		X
849	showjumping	N	C+D	41	COMP				X
850	cytoplasm	N	C	126	COMP				
851	hissing	N	D	47	COMP				
852	craftsmanship	N	C	188	COMP			X	
853	nation	N	D	8431	COMP	X			
854	microorganism	N	C	155	COMP				
855	pigmentation	N	D	54	COMP				
856	repugnance	N	D	36	COMP	X	X		
857	bodyline	N	C	33	COMP			X	
858	accusation	N	D	1121	COMP				
859	breadth	N	D	574	COMP	X			
860	houseplant	N	C	64	COMP				
861	affluence	N	D	207	COMP	X			
862	manipulator	N	D	65	COMP				
863	tuesday	N	C	3609	COMP	X	X		
864	generalisation	N	D	312	COMP				
865	carpark	N	C	39	COMP				
866	interview	N	D	6516	COMP				X
867	elasticity	N	D	386	COMP				
868	dosage	N	D	148	COMP	X			
869	conqueror	N	D	255	COMP				
870	consulting	N	D	144	COMP				
871	aphorism	N	D	79	COMP	X	X		
872	review	N	D	8919	COMP			X	
873	zeolite	N	C	86	COMP				
874	paperweight	N	C	60	COMP				
875	dynamism	N	D	144	COMP	X			
876	milking	N	D	161	COMP				
877	evolutionist	N	D	75	COMP				
878	disarmament	N	D	426	COMP	X	X		
879	ruthlessness	N	D	110	COMP				

880	enormity	N	D	128	COMP	X			
881	shrinkage	N	D	119	COMP	X			
882	child-care	N	C	88	COMP				
883	jewellery	N	D	1230	COMP				
884	baby-sitter	N	C+D	37	COMP				
885	pressure	N	D	14415	COMP	X			
886	taming	N	D	31	COMP				
887	humanist	N	D	92	COMP				
888	bluebird	N	C	68	COMP				X
889	urinal	N	D	60	COMP	X			X
890	check-in	N	C	48	COMP		X		X
891	mohair	N	C	50	COMP		X		
892	free-kick	N	C	130	COMP				X
893	separatism	N	D	134	COMP				
894	classicism	N	D	88	COMP				
895	estimator	N	D	85	COMP				
896	entrant	N	D	609	COMP	X			
897	washing-machine	N	C+D	36	COMP				X
898	amenity	N	D	781	COMP	X	X		
899	nursery	N	D	2060	COMP				X
900	biochemistry	N	C+D	241	COMP				
901	funfair	N	C	49	COMP				X
902	carpenter	N	D	623	COMP	X	X		
903	prosecutor	N	D	567	COMP				
904	remuneration	N	D	487	COMP				
905	underneath	N	C	33	COMP		X		
906	policing	N	D	606	COMP				
907	rerun	N	D	33	COMP				
908	hoving	N	D	33	COMP				
909	scholarship	N	D	1098	COMP			X	
910	licencee	N	D	35	COMP				
911	dimity	N	D	89	COMP	X	X		
912	resumption	N	D	310	COMP	X			
913	nightwatchman	N	C	30	COMP				
914	flip-flop	N	C	63	COMP				X
915	theorizing	N	D	74	COMP				
916	structuring	N	D	179	COMP				
917	firework	N	C	508	COMP				X
918	inservice	N	C	88	COMP				X
919	opportunism	N	D	99	COMP				X
920	proficiency	N	D	174	COMP	X			
921	enhancement	N	D	603	COMP	X			
922	starfish	N	C	102	COMP		X		X
923	allotment	N	D	376	COMP	X		X	
924	superiority	N	D	754	COMP				
925	vengeance	N	D	372	COMP	X			
926	bedchamber	N	C	62	COMP				
927	tigress	N	D	49	COMP				
928	disparity	N	D	397	COMP				
929	uptake	N	C	359	COMP				X
930	soliton	N	D	31	COMP				
931	break-out	N	C	36	COMP				X
932	casework	N	C	101	COMP				X
933	binding	N	D	421	COMP				
934	walking-stick	N	C+D	35	COMP				
935	inductivist	N	D	94	COMP				
936	backlog	N	C	216	COMP			X	
937	ribber	N	D	218	COMP				
938	package	N	D	7137	COMP	X			

939	couplet	N	D	147	COMP				X
940	ice-cap	N	C	32	COMP				X
941	optimism	N	D	837	COMP	X			X
942	repository	N	D	337	COMP	X			
943	putter	N	D	191	COMP				
944	unrest	N	D	913	COMP				X
945	negotiator	N	D	491	COMP				
946	anaesthetist	N	D	70	COMP				
947	fecundity	N	D	76	COMP				
948	physicist	N	D	544	COMP	X			
949	insecticide	N	C	172	COMP				
950	blocking	N	D	98	COMP				
951	blackbird	N	C	300	COMP				X
952	schoolmaster	N	C	302	COMP				X
953	depress	V	D	350	COMP				X
954	reinforce	V	D	2749	COMP				X
955	pressure	V	D	149	COMP	X			
956	venerate	V	D	49	COMP	X	X		
957	beware	V	D	558	COMP	X	X		
958	localise	V	D	104	COMP				
959	reawaken	V	D	48	COMP				
960	overhaul	V	C	188	COMP		X		X
961	exterminate	V	D	80	COMP				X
962	recede	V	D	448	COMP	X	X		
963	over-estimate	V	C	60	COMP				
964	out-perform	V	C	31	COMP				
965	dismember	V	D	88	COMP				X
966	transpose	V	D	163	COMP	X	X		
967	migrate	V	D	635	COMP	X			
968	surmount	V	D	255	COMP				X
969	placate	V	D	165	COMP	X	X		
970	revolve	V	D	490	COMP	X	X		
971	pin-point	V	C	57	COMP				X
972	encompass	V	D	868	COMP	X			X
973	vacillate	V	D	34	COMP	X	X		
974	mortify	V	D	64	COMP	X			X
975	colonise	V	D	186	COMP				
976	animate	V	D	146	COMP	X	X		
977	prostrate	V	D	33	COMP	X	X		
978	interface	V	D	265	COMP		X		X
979	rethink	V	D	185	COMP				
980	reconstitute	V	D	194	COMP			X	
981	suffocate	V	D	208	COMP	X	X		
982	relegate	V	D	380	COMP	X	X		
983	publicise	V	D	419	COMP	X			
984	rehabilitate	V	D	158	COMP	X			
985	telescope	V	C	34	COMP				
986	reinvest	V	D	59	COMP				
987	adjudicate	V	D	121	COMP	X			X
988	commentate	V	D	33	COMP		X		
989	individuate	V	D	35	COMP	X			
990	engineer	V	D	310	COMP	X		X	
991	revitalise	V	D	163	COMP				
992	unearth	V	D	263	COMP			X	
993	forward	V	D	294	COMP	X			
994	sweeten	V	D	126	COMP	X			X
995	untangle	V	D	46	COMP				
996	attune	V	D	126	COMP	X			
997	reactivate	V	D	84	COMP				

998	surface	V	D	468	COMP	X			
999	maximise	V	D	719	COMP	X			
1000	specialise	V	D	1162	COMP				

Part II – The OED sample

Number	Lexeme	WF process	WC	SA 4	SA 3	FA	CA	Semantic subtype	Field -specific	Particle compound	Evaluative
1	overall	C	ADJ	X		ph					
2	online	C	ADJ	X				metonymy			
3	full time	C	ADJ	X				specialization			
4	straightforward	C	ADJ	X		acf	X	metaphor			
5	overseas	C	ADJ	X		sa	X				
6	part-time	C	ADJ	X				specialization			
7	way-out	C	ADJ	X		acf	X				E
8	present-day	C	ADJ		X			generalization			
9	way back	C	ADJ		X						
10	left-behind	C	ADJ	X				specialization		P	
11	day-to-day	C	ADJ	X		ph					
12	no-good	C	ADJ	X		ph					E
13	short-run	C	ADJ	X				metaphor			
14	long-range	C	ADJ		X			specialization			
15	high-level	C	ADJ	X				specialization			
16	high-pressure	C	ADJ	X		as		metaphor			
17	go-ahead	C	ADJ	X				metaphor		P	
18	left-out	C	ADJ		X			metaphor		p	
19	white-collar	C	ADJ	X				metonymy			
20	face-to-face	C	ADJ	X		ph		metonymy			
21	hung-up	C	ADJ	X						P	
22	one-one	C	ADJ	X		acf	X				
23	embedded	C	ADJ		X			metaphor			
24	far-back	C	ADJ		X						
25	automotive	C	ADJ	X							
26	full-scale	C	ADJ		X			metaphor			
27	pluralistic	C	ADJ	X							
28	second-class	C	ADJ		X			generalization			
29	three-way	C	ADJ		X	ph		metaphor / metonymy			
30	used-up	C	ADJ		X			generalization		P	E
31	age-old	C	ADJ	X			t				
32	low-grade	C	ADJ		X			metaphor			E
33	overdue	C	ADJ	X		sa					
34	double-blind	C	ADJ	X			X	metaphor	T		
35	high-grade	C	ADJ		X			metaphor			E
36	undercover	C	ADJ		X			metaphor			
37	upfront	C	ADJ	X							
38	far-out	C	ADJ		X			metaphor			E
39	front-end	C	ADJ	X							

40	straight-out	C	ADJ	X				amelioration			E
41	someday	C	ADV	X		ph		generalization			
42	on-stage	C	ADV		X	ph		metonymy			
43	must-be	C	N	X		ph		exocentric			E
44	output	C	N	X				specialization		P	
45	set-up	C	N	X						P	
46	database	C	N	X							
47	feedback	C	N	X				metaphor		P	
48	breakdown	C	N	X				metaphor		P	
49	make-up	C	N	X						P	
50	carry-out	C	N	X				specialization		P	
51	set-out	C	N	X						P	
52	wavelength	C	N	X				specialization	T		
53	viewpoint	C	N		X			metaphor			
54	come-back	C	N	X				specialization		P	
55	give-up	C	N	X						P	
56	airport	C	N	X			X	metaphor / metonymy			
57	bureaucracy	C	N	X							
58	get-out	C	N	X				specialization		P	
59	labour force	C	N	X				metonymy	T		
60	lay-out	C	N	X				specialization		P	
61	trade union	C	N	X				specialization	T		
62	backup	C	N	X						P	
63	standpoint	C	N		X			metaphor			
64	take-up	C	N	X						P	
65	side effect	C	N	X				metaphor			
66	third party	C	N	X				specialization	T		
67	uptake	C	N	X				specialization		P	
68	airplane	C	N	X			X				
69	build-up	C	N	X				metaphor		P	
70	put-up	C	N	X						P	
71	airline	C	N		X			metonymy			
72	overlap	C	N		X			metaphor		P	
73	manpower	C	N		X			metonymy			
74	set-aside	C	N	X				specialization	T	P	
75	wild life	C	N	X				specialization			
76	get-up	C	N	X				exocentric		P	
77	trademark	C	N	X				specialization	T		
78	deadline	C	N	X			X				
79	workforce	C	N	X				metonymy	T		
80	basketball	C	N	X				exocentric			
81	biomass	C	N		X			specialization			
82	social work	C	N	X				specialization			
83	breakthrough	C	N	X				metaphor		P	
84	folklore	C	N	X		acf	t				
85	put-down	C	N	X				exocentric		P	
86	update	C	N	X						P	

87	worksheet	C	N		X			specialization			
88	natural gas	C	N	X				specialization	T		
89	hallway	C	N		X						
90	website	C	N	X				metaphor	T		
91	check-list	C	N		X						
92	go-away	C	N	X				exocentric			
93	drop-out	C	N	X				exocentric		P	E
94	girlfriend	C	N		X			specialization			
95	waveform	C	N		X			specialization	T		
96	inflow	C	N		X			metaphor / metonymy		P	
97	show-up	C	N	X						P	
98	slow-down	C	N		X			specialization		P	
99	cardboard	C	N	X				specialization			
100	shut-down	C	N	X						P	
101	boyfriend	C	N		X			specialization			
102	stock-market	C	N	X				specialization	T		
103	by-pass	C	N	X				specialization	T	P	
104	cut-out	C	N		X					P	
105	takeover	C	N	X				specialization		P	
106	heart attack	C	N	X				metaphor			
107	suitcase	C	N	X				generalization			
108	trade-off	C	N	X						P	
109	desk-top	C	N	X				exocentric			
110	machine gun	C	N	X			t		T		
111	motion picture	C	N	X				specialization			
112	time series	C	N	X					T		
113	back-down	C	N	X				exocentric		P	
114	workstation	C	N	X				specialization			
115	one-step	C	N	X		ph	X	metonymy			
116	border-line	C	N		X			metaphor			
117	paperback	C	N	X				exocentric			
118	pickup	C	N	X						P	
119	spreadsheet	C	N	X				specialization	T		
120	foodstuff	C	N	X		acf	t				
121	standout	C	N	X				exocentric/amelior./metaphor		P	E
122	air force	C	N	X				metonymy	T		
123	cutback	C	N	X						P	
124	driveway	C	N	X			t	specialization			
125	eigenvalue	C	N	X		lc	X		T		
126	run-out	C	N		X			specialization	T	P	
127	artwork	C	N		X			specialization			
128	benchmark	C	N	X				metaphor			
129	science fiction	C	N	X				specialization			
130	blueprint	C	N		X			metaphor			
131	clean-up	C	N		X			metaphor		P	
132	runway	C	N	X							

133	workload	C	N	X				metaphor			
134	come-over	C	N	X				exocentric		P	
135	spread-over	C	N		X					P	
136	mass medium	C	N	X				specialization	T		
137	overtone	C	N		X			metaphor			
138	pay-off	C	N	X				specialization		P	
139	shot-gun	C	N	X			t		T		
140	up-grade	C	N		X					P	
141	ceasefire	C	N	X		acf	X	specialization			
142	close-up	C	N	X		acf	X	exocentric			
143	let-out	C	N	X				specialization		P	
144	mass media	C	N	X				specialization	T		
145	role model	C	N	X				specialization			
146	throughput	C	N	X				specialization		P	
147	world war	C	N	X					T		
148	airfield	C	N	X			X	metonymy			
149	bed-rock	C	N	X			X	metaphor			
150	mutual fund	C	N	X				specialization	T		
151	speed-up	C	N		X			specialization		P	
152	write-down	C	N	X				specialization	T	P	
153	bad news	C	N		X			metonymy			E
154	go-off	C	N	X						P	
155	heartbeat	C	N		X			metaphor			
156	know-nothing	C	N	X		ph		exocentric			E
157	turn-off	C	N		X			exocentric		P	
158	astronaut	C	N	X				metonymy			
159	seafood	C	N	X							
160	snap-shot	C	N		X			metaphor			
161	time frame	C	N	X				metaphor			
162	uplift	C	N		X					P	
163	get-together	C	N	X		ph		specialization			
164	hard disk	C	N	X				specialization	T		
165	hold-up	C	N	X						P	
166	T-shirt	C	N	X		ac	X				
167	underwear	C	N	X				specialization			
168	wild type	C	N	X				specialization	T		
169	horsepower	C	N	X				metaphor	T		
170	mark-up	C	N	X						P	
171	spotlight	C	N		X			metaphor			

172	back-out	C	N	X						P	
173	cell phone	C	N	X					T		
174	come-along	C	N	X						P	
175	cover-up	C	N	X						P	
176	feel-good	C	N	X				metonymy / exocentric			
177	hold-out	C	N	X				exocentric		P	
178	leftover	C	N	X				exocentric		P	
179	line-up	C	N	X						P	
180	outcrop	C	N		X			metaphor / specialization		P	
181	pull-out	C	N	X				exocentric		P	
182	run-off	C	N	X				exocentric		P	
183	door-step	C	N		X			metaphor			
184	dug-out	C	N	X				exocentric		P	
185	freelance	C	N	X			X	exocentric			
186	freeway	C	N		X				T		
187	hind-sight	C	N	X		lc	X				
188	landfill	C	N	X				exocentric			
189	pressure group	C	N	X				specialization	T		
190	sort-out	C	N	X						P	
191	timeout	C	N	X					T		
192	top level	C	N		X			metonymy			
193	workout	C	N	X						P	
194	back seat	C	N		X			metaphor			
195	go-around	C	N		X					P	
196	grandparent	C	N	X							
197	headlight	C	N		X			metaphor			
198	outreach	C	N	X				metaphor		P	
199	printout	C	N	X				exocentric		P	
200	give-away	C	N		X			metaphor / metonymy		P	
201	hard drive	C	N	X					T		
202	mirror image	C	N		X			metaphor			
203	must-see	C	N	X		ph					E
204	nightclub	C	N	X				specialization			
205	no-account	C	N	X		ph					E
206	overcoat	C	N	X				specialization			
207	screenplay	C	N	X							
208	short-fall	C	N	X							
209	shut-off	C	N	X				exocentric		P	
210	warm-up	C	N	X				metaphor / exocentric		P	

211	windshield	C	N	X							
212	back-lash	C	N		X			metaphor			
213	breakaway	C	N	X						P	
214	food chain	C	N	X				metaphor			
215	hot spot	C	N		X			metaphor			
216	payload	C	N	X				specialization			
217	put-away	C	N	X				exocentric	T	P	
218	run-up	C	N	X						P	
219	upsurge	C	N	X			t				
220	focus group	C	N		X			specialization	T		
221	handout	C	N	X				exocentric		P	
222	laptop	C	N	X				exocentric			
223	rainforest	C	N	X			X	metonymy			
224	read-out	C	N	X				specialization		P	
225	turnaround	C	N		X			metaphor / amelioration		P	
226	aeroplane	C	N	X							
227	dead end	C	N	X			X	metaphor			
228	fast food	C	N	X			X	metonymy			
229	heavy-weight	C	N	X				exocentric	T		E
230	high-rise	C	N	X				exocentric			
231	knock-out	C	N	X						P	
232	playback	C	N	X						P	
233	shut-out	C	N	X						P	
234	spillover	C	N		X			metaphor		P	
235	step-up	C	N	X				metaphor		P	
236	wipe-out	C	N		X			metaphor		P	
237	blow-up	C	N	X						P	
238	body language	C	N	X				metaphor			
239	go-round	C	N	X						P	
240	hangover	C	N	X				exocentric		P	
241	home page	C	N		X				T		
242	home run	C	N		X				T		
243	look-over	C	N	X				exocentric		P	
244	medicine man	C	N	X					T		
245	round trip	C	N		X			generalization			
246	run-over	C	N	X						P	
247	soap opera	C	N	X			X	metaphor / metonymy			
248	spokesperson	C	N	X		ac	X				
249	war-lord	C	N	X				specialization			

250	wasteland	C	N	X							E
251	billboard	C	N	X							
252	black box	C	N	X							
253	break-off	C	N		X			metaphor		P	
254	carry-over	C	N	X				exocentric		P	
255	handbag	C	N		X			specialization / generalization			
256	open system	C	N		X			specialization	T		
257	quarterback	C	N	X				exocentric	T		
258	skyline	C	N		X			metonymy / specialization			
259	stir-up	C	N		X			metaphor		P	
260	walk-up	C	N	X				exocentric		P	
261	big brother	C	N	X				metaphor / deterioration			E
262	flashback	C	N		X			exocentric		P	
263	higher-up	C	N	X		acf	t	exocentric			E
264	spin-off	C	N	X				exocentric		P	
265	timeline	C	N		X						
266	call-up	C	N	X				exocentric		P	
267	upkeep	C	N	X						P	
268	buy-out	C	N	X						P	
269	chairperson	C	N	X			X	metonymy			
270	pull-up	C	N	X				exocentric		P	
271	break-even	C	N	X					T	P	
272	highlight	C	V	X			X				
273	update	C	V	X						P	
274	broadcast	C	V		X			metaphor			
275	by-pass	C	V		X			metaphor			
276	upgrade	C	V		X					P	
277	network	C	V		X			metaphor			
278	download	C	V	X						P	
279	layer	C	V	X		as	X				
280	streamline	C	V	X				metaphor			
281	problem-solve	C	V	X			t				
282	pinpoint	C	V	X				metaphor			
283	outsource	C	V	X				specialization		P	
284	ongoing	C+D	ADJ	X						P	
285	handicapped	C+D	ADJ	X		ac	X				
286	far-reaching	C+D	ADJ		X			metaphor			
287	two-dimensional	C+D	ADJ		X			metaphor			E
288	cross-sectional	C+D	ADJ		X			metaphor			

289	open-ended	C+D	ADJ		X			metaphor			
290	overriding	C+D	ADJ	X				metaphor		P	
291	mixed-up	C+D	ADJ		X			metaphor			
292	one-dimensional	C+D	ADJ		X			metaphor / deterioration			E
293	stereotyped	C+D	ADJ	X				metaphor			E
294	stereotypical	C+D	ADJ	X							
295	self-defeating	C+D	ADJ		X			metaphor			
296	withdrawal	C+D	N	X			X				
297	shareholder	C+D	N	X				specialization	T		
298	cross section	C+D	N		X			metaphor	T		
299	outsider	C+D	N	X		acf	X	metaphor			
300	broadcasting	C+D	N		X			metaphor			
301	telecommunicatio n	C+D	N	X				specialization			
302	family planning	C+D	N		X			specialization			
303	turning-point	C+D	N		X			metaphor			
304	upheaval	C+D	N		X			metaphor		P	
305	data processing	C+D	N		X			specialization	T		
306	insider	C+D	N	X		acf	X	metaphor			
307	value-added	C+D	N	X		sa	X	specialization	T		
308	human resources	C+D	N	X				metonymy			
309	concentration camp	C+D	N	X				euphemism			
310	networking	C+D	N	X				specialization	T		
311	care-taker	C+D	N		X			specialization			
312	air conditioning	C+D	N	X				specialization			
313	information technology	C+D	N	X				specialization	T		
314	pacemaker	C+D	N		X			metaphor			
315	broadcaster	C+D	N		X			metaphor			
316	outgrowth	C+D	N	X				metaphor / metonymy			
317	high technology	C+D	N	X				metaphor			
318	word processing	C+D	N	X				metonymy			
319	fundraising	C+D	N	X			X	metaphor			
320	loudspeaker	C+D	N	X							
321	cutting edge	C+D	N	X				metaphor			
322	homemaker	C+D	N	X				specialization			
323	word processor	C+D	N	X				metonymy			
324	bestseller	C+D	N	X							E
325	bread-winner	C+D	N	X							
326	smallholder	C+D	N	X							
327	air conditioner	C+D	N	X							

328	basic	D	ADJ		X			metaphor			
329	environmental	D	ADJ		X			specialization			
330	nuclear	D	ADJ		X			specialization	T		
331	developing	D	ADJ		X			euphemism			E
332	operational	D	ADJ		X						
333	exceptional	D	ADJ		X			amelioration			E
334	systemic	D	ADJ		X				T		
335	demanding	D	ADJ	X				specialization			E
336	distal	D	ADJ	X				specialization	T		
337	viral	D	ADJ		X			metaphor			
338	colourful	D	ADJ		X			metaphor			E
339	infra-red	D	ADJ	X			X		T		
340	post-war	D	ADJ		X			specialization			
341	knowledgeable	D	ADJ	X		acf	X				E
342	fluorescent	D	ADJ		X	acf	X	determinization	T		
343	bipolar	D	ADJ		X			metaphor			
344	ultraviolet	D	ADJ	X		acf	X		T		
345	phenomenal	D	ADJ		X						E
346	institutionalized	D	ADJ		X			specialization			E
347	multimedia	D	ADJ		X			specialization			
348	pivotal	D	ADJ		X			metaphor			E
349	competing	D	ADJ		X						
350	expatriate	D	ADJ	X							
351	sensational	D	ADJ		X						
352	cooked	D	ADJ		X			metaphor			E
353	caring	D	ADJ		X			specialization			
354	polarized	D	ADJ		X			metaphor			
355	probabilistic	D	ADJ		X	ac			T		
356	processed	D	ADJ		X			specialization	T		
357	inter-war	D	ADJ		X			specialization			
358	multidimensional	D	ADJ		X			metaphor			
359	pre-war	D	ADJ		X			specialization			
360	correctional	D	ADJ	X				metonymy			
361	narcissistic	D	ADJ	X				commonization			E
362	opportunistic	D	ADJ	X				deterioration			E
363	exhausting	D	ADJ		X			hyperbole			E
364	telling	D	ADJ	X				metaphor			E
365	resourceful	D	ADJ		X			amelioration			E
366	preterm	D	ADJ		X			specialization	T		

367	fluffy	D	ADJ		X			metaphor			
368	sketchy	D	ADJ		X			metaphor / deterioration			E
369	tailored	D	ADJ	X		acf	X				
370	exceptionally	D	ADV		X						E
371	chronically	D	ADV		X			generalization			
372	exponentially	D	ADV	X				determinization			
373	movie	D	N	X		acf	X				
374	interface	D	N	X			X				
375	Internet	D	N	X							
376	coverage	D	N		X			specialization			
377	processor	D	N		X			specialization	T		
378	shortage	D	N	X							
379	infrastructure	D	N	X				specialization			
380	reactor	D	N		X				T		
381	activist	D	N	X				specialization			
382	inhibitor	D	N		X			specialization	T		
383	literacy	D	N	X				specialization			
384	seasonal	D	N	X		as					
385	buffer	D	N	X		lc	X				
386	catalyst	D	N	X		ac	X		T		
387	mutant	D	N		X	as					E
388	detective	D	N	X		as		specialization / exocentric			
389	dosage	D	N	X			t				
390	mobilization	D	N	X				specialization			
391	platelet	D	N	X				specialization	T		
392	documentary	D	N	X		as		exocentric			
393	dismissal	D	N	X							
394	interviewer	D	N	X			X				
395	enactment	D	N	X				specialization	T		
396	adrenal	D	N	X				exocentric	T		
397	organizer	D	N		X						
398	capacitor	D	N	X		acf	X		T		
399	globalization	D	N	X					T		
400	microwave	D	N	X				exocentric			
401	privatization	D	N		X			specialization	T		
402	monograph	D	N	X				specialization	T		
403	sewage	D	N	X		lc	X				
404	sensor	D	N	X				specialization	T		
405	wireless	D	N	X		as		exocentric			

406	steamer	D	N	X		acf	X				
407	supermarket	D	N	X				specialization	T		
408	booklet	D	N		X			specialization			
409	embodiment	D	N		X			metaphor			
410	empathy	D	N	X							
411	functionality	D	N		X						
412	ultrasound	D	N	X			X		T		
413	deductible	D	N	X		as		exocentric	T		
414	capacitance	D	N	X		acf	t		T		
415	activism	D	N	X							
416	collaborator	D	N		X			deterioration			E
417	decentralization	D	N	X				specialization			
418	interstate	D	N	X				exocentric			
419	instrumentation	D	N		X			metonymy			
420	relativity	D	N		X			specialization	T		
421	router	D	N	X				specialization	T		
422	rationalization	D	N		X			specialization			
423	subroutine	D	N		X			specialization	T		
424	leftist	D	N	X			X	specialization	T		
425	deterrence	D	N		X			specialization	T		
426	superpower	D	N		X			metonymy			
427	championship	D	N	X				metonymy			
428	recombinant	D	N	X				exocentric			
429	conductance	D	N	X				specialization	T		
430	extremist	D	N		X			specialization / deterioration			E
431	grader	D	N		X	acf	X				
432	normalization	D	N		X			specialization	T		
433	positioning	D	N	X		acf	t				
434	subway	D	N	X				specialization			
435	foreseeable	D	N	X							
436	floppy	D	N	X				exocentric			
437	packaging	D	N	X			t				
438	shrinkage	D	N		X			metaphor			
439	interplay	D	N	X			X				
440	tanker	D	N	X		acf	X				
441	extrapolation	D	N	X		ac	X				
442	vegetarian	D	N	X		ac					
443	columnist	D	N	X				specialization	T		
444	magnetization	D	N		X			metaphor			

445	populist	D	N	X					T		E
446	connectivity	D	N		X						
447	dumping	D	N		X			metaphor	T		
448	dynamism	D	N		X						
449	footage	D	N	X			X	metonymy			
450	institutionalizatio n	D	N		X			specialization			
451	internet	D	N	X				metaphor	T		
452	revisionist	D	N	X				specialization	T		
453	positivism	D	N	X					T		
454	poster	D	N		X			generalization			
455	surround	D	N	X				exocentric			
456	fractionation	D	N	X		acf	t				
457	hippie	D	N	X							
458	diner	D	N		X			metonymy			
459	weighting	D	N		X			specialization / metaphor / amelior.			
460	retiree	D	N	X		acf	X				
461	blender	D	N		X						
462	fledgeling	D	N	X		acf	X				
463	foreword	D	N	X				metonymy / specialization			
464	subtitle	D	N	X							
465	flotation	D	N	X		ac	X	specialization	T		
466	consumerism	D	N	X		acf	t	specialization			
467	detainee	D	N	X				specialization			
468	opportunist	D	N	X				deterioration			E
469	tabloid	D	N	X				specialization	T		
470	foreground	D	N		X			metaphor			
471	formulate	D	V		X						
472	interview	D	V	X							
473	mobilize	D	V	X				specialization			
474	recycle	D	V		X			specialization			
475	package	D	V		X	as		metaphor			
476	polarize	D	V		X			metaphor			
477	encapsulate	D	V		X			metaphor			
478	institutionalize	D	V		X						
479	interface	D	V	X			X				
480	initialize	D	V	X				specialization	T		
481	sensitize	D	V		X		X	generalization			
482	buffer	D	V	X		lc	X				
483	customize	D	V	X							

484	understate	D	V	X			X				
485	resource	D	V	X							
486	contraceptive	other	ADJ	X		ac	X				
487	mechanistic	other	ADJ	X		ac	X	deterioration			E
488	redox	other	ADJ	X		ac	X		T		
489	pixel	other	N	X		ac	X		T		
490	rotor	other	N	X		ac	X				
491	cultivar	other	N	X		ac	X				
492	surfactant	other	N	X		ac	X				
493	aerosol	other	N	X		ac		specialization	T		
494	soccer	other	N	X		ac	X		T		
495	contraception	other	N	X		ac	X				
496	motel	other	N	X		ac	X				
497	high-tech	other	N	X		ac	X				
498	catalyse	other	V	X		ac	X		T		
499	extrapolate	other	V	X		ac	X				
500	adsorb	other	V	X		ac	X				