



Kateřina Vašků

Lexical idioms in English

Lexikální idiomy v angličtině

Teze

Vedoucí práce: Prof. PhDr. Aleš Klégr 2018 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. Describing such items as idioms is not generally accepted yet, inasmuch as idioms are canonically thought to be multi-word expressions (cf. Howarth 1998, Burger 1998, etc.). The study is an attempt to argue for the possibility of describing and analysing lexical idioms by similar criteria as idiomatic multi-word expressions.

Chapter 2 provides an overview of the traditional and more recent approaches to phraseology, discusses the demarcations between phraseology and other disciplines arguing that phraseology overlaps with several disciplines, especially morphology, syntax and semantics. The chapter also presents structural classifications of phrasemes in the work of some of the most influential phraseologists within the phraseological approach, focusing on the British (Cowie and Howarth) and Continental (Burger, Čermák) tradition. The classifications of phraseological units by these and other influential authors show that apart from Čermák, it is generally taken for granted that phraseology operates on the level of word combinations only. Finally, the chapter describes the main concepts related to semantic aspects of idioms, especially (non-)compositionality, idiomaticity and anomaly attempting to establish the theoretical background for idiom identification.

Chapter 3 summarizes references in the literature to idiomaticity on the lexical level, 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. The chapter first presents the theory of semantic coindexation, which accounts for the way meaning construction occurs in derivation and compounding. The following two parts focus on scattered references to idiomatic derivatives and compounds in literature. The aim of this heterogeneous list of mentions is to demonstrate that these references to phraseological derivative and (especially) compounds are not scarce as one might expect. The chapter also two lines of linguistics which are closely related to the topic of lexical idioms: the cognitive approach represented by Benczes (2005, 2006 & 2015) and the account of meaning-predictability from the onomasiological perspective (Dokulil, 1978 & Štekauer, 2005a, b). Although these approaches do not deal explicitly with the inclusion of lexical idioms into phraseology, their subject of study is in fact very much related.

Chapter 4 is the last chapter of the theoretical part and introduces probably the only systematic treatment of lexical idioms within the discipline of phraseology, i.e. the theory of

²

lexical idioms by Čermák (2007a, b), whose theoretical work serves as the basis for this study. The main section of the chapter presents Čermák's morphological classification of lexical phrasemes (Čermák uses the term phraseme when the form of the unit is in focus and the term idiom when pointing to its meaning). Lexical phrasemes are divided according to the word-formation process involved into derivational and compositional phrasemes and these types are further divided according to the word-class of the lexeme. Chapter 4 also presents a quantitative analysis of lexical idioms in Czech by Klötzerová (1997, 1998), who tests and confirms Čermák's theory in practice.

Chapter 5 describes objectives of the research and poses research questions to be answered by the study and presents the provisional definition of lexical idioms. In addition, it provides information about the data and their processing. The research questions posed in this study are as follows:

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

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?

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?

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

The provisional definition formulated before the analyses describes lexical idioms as singleword lexemes formed as combinations of components which are anomalous semantically and/or collocationally and/or grammatically. **Chapter 6** is divided into two stages and presents two analyses, each with its own objectives. The aim of the first stage is 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 obtained in the first stage, a new definition is formulated, which is used to retrieve a new sample of items meeting the criteria for lexical idioms.

The first sample consists of 1 000 randomized lexemes retrieved from the BNC. The lexemes are classified into several categories reflecting the presence or absence of anomalous features to find out which irregularities should be included among the criteria for lexical idioms. The features taken into account involve: simple/complex structure, word-class and word-formation process involved. The categories of complex lexemes are as follows:

a) Category 0: complex lexemes with formally and semantically regular structure;

b) Category 1: formally anomalous complex lexemes;

c) Category 2: collocationally anomalous complex lexemes;

d) Category 3: semantically anomalous complex lexemes which have both regular and idiomatic senses;

e) Category 4: semantically anomalous complex lexemes with idiomatic meaning only.

The following tables quantify the distribution of word-classes in the sample and the categories described above.

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

Table 1: Distribution of word-classes and simple/complex lexemes in the BNC sample

BNC SAMPLE	ТҮРЕ	TOTAL OF	TOTAL OF ITEMS
COMPLEX LEXEMES		LEXEMES IN THE SAMPLE	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 2: Categories of complex lexemes in the BNC sample

The combination of word-class/word-formation process data and the assigned category are presented below:

WC	CL	%	CAT 0	%	CAT 1	%	CAT 2	%	CAT 3	%	CAT 4	%
Ν	322	47.3	158	41.3	74	46.8	49	60.5	21	51.2	86	67.2
ADJ	264	38.8	171	45.0	56	35.4	15	18.5	17	41.5	26	20.3
ADV	47	6.9	40	10.5	3	1.9	3	3.7	0	0	4	3.1
V	48	7.0	12	3.2	25	15.9	14	17.3	3	7.3	12	9.4
Т	681	100.0	381	100.0	158	100.0	81	100.0	41	100.0	128	100.0

Table 3: Word-class distribution in all categories of complex lexemes in the BNC sample

WF	CL	%	CAT 0	%	CAT 1	%	CAT 2	%	CAT 3	%	CAT 4	%
D	523	76.8	311	81.6	146	92.4	51	63	30	73.1	60	46.9
С	119	17.5	40	10.5	12	7.6	29	35.8	9	22.0	61	47.6
D+C	39	5.7	30	7.9	0	0	1	1.2	2	4.9	7	5.5
Т	681	100.0	381	100.0	158	100.0	81	100.0	41	100.0	128	100

Table 4: Distribution of word-formation processes in all categories of complex lexemes in the BNC sample

The problematic subtypes of anomalous lexemes discussed in the thesis involve:

a) Latin or Greek borrowings or English formations based on these source languages with an unproductive affix and/or bound base (126 lexemes of 158 assigned to Category 1): *atonement, dilatory, reversal, scripture, literate, scientific, pressure*

b) Latin or Greek borrowings with a component used systematically in English and another component with no systematic use in English: *garrulous, duplicate, cadence, tirade, relegate*

c) lexemes of Category 3 (mostly idiomatic in a shifted sense): *mouth-watering, tubby, offshoot, goer, unearth, oddity, skinny, protector, tipper*

Other, more prototypical instances of idioms are presented below:

Category 1: verbs: forward, engineer, pressure; elderly, olden, topmost, father-in-law, stompie, showbiz

Category 2: mohair, Tuesday, mulberry, mistletoe, best, registrar, carpenter

The BNC sample analysis shows 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. It seems more reasonable 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 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.

6

The amended definition was then applied when collecting the OED sample of 500 lexical idioms for the second analysis. The sample was then analysed in terms of the anomaly involved and then an attempt was made to identify a possible subtype of the anomaly found.

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		

The following table summarizes the most distinct semantic subtypes:

Table 5: Distribution of semantic subtypes within the OED sample

The subtypes are exemplified below:

SEMANTIC	EXAMPLES OF LEXEMES
SUBTYPES	
specialization	output, get-out, let-out, mobilization, activist, booklet, post-war
metaphor	multidimensional, encapsulate, viewpoint, blueprint, work load
exocentricity	know-nothing, paperback, close-up, high-rise, print-out, detective,
	documentary, floppy, microwave
metonymy	white-collar, face-to-face, footage, foreword, outgrowth
generalization	suitcase, poster, used-up, handbag
deterioration	collaborator, mechanistic, opportunist
amelioration	exceptional, resourceful, standout
no specific subtype	make-up, back-up, hallway, grandparent, screenplay, black box,
	windshield

Table 6: Examples of semantic subtypes within the OED sample

The following tables provide data concerning the formal and collocational anomaly and their subtypes and illustrate the subtypes with examples:

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

Table 7: Formal and collocational anomaly in the OED sample

SUBCATEGORY OF FORMAL AND	EXAMPLES OF LEXEMES
COLLOCATIONAL ANOMALY	
anomalous component form	contraceptive, cultivar, catalyse, spokesperson,
	handicapped
anomalous suffix	nouns: documentary, wireless, deductible;
	verbs: package, layer; adjective: high-pressure
other formal anomalies	no-good, must-see, value-added, overdue,
	overseas
formally anomalous combination	movie, steamer, capacitor, knowledgeable,
	tailored
tautology	packaging, dosage, capacitance, fractionation,
	shot-gun, machine gun, driveway
semantic incompatibility	deadline, dead end, fast food, interface,
	airport, soap opera
low collocability	buffer, sewage, eigenvalue, hind-sight

As far as the formal structure of lexical idioms is concerned, the category of lexical idioms in English divides into four main groups in terms of the word-formation process involved: compounds, derivatives, formations combining both composition and derivation and other formations (such as clipping compounds, blends, etc.). 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.

STRUCTURAL	NUMBER OF	%	EXAMPLES
ТҮРЕ	LEXEMES	, ,	
DERIVATIVES	158	31.6	processor, reactor, organizer, grader, tanker; distal, inter-war, institutionalized
COMPOUNDS	283	56.6	viewpoint, horsepower, natural gas, double-blind, overall, set-up, break-down
COMBINED FORMATIONS	45	9	pacemaker, loudspeaker, bread-winner, open-ended, two-dimensional, word processing, cross-sectional
OTHER FORMATIONS	14	2.8	contraceptive, cultivar, surfactant, catalyse
TOTAL	500	100	

The categories are quantified and exemplified below:

Table 8: Structural subtypes of lexical idioms and their examples

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).

As far as the degree of idiomaticity for lexical idioms is concerned, it seems that there are at least three least three scalar axes or variables of idiomaticity. They are illustrated in table below:

	axes/variables of idiomaticity	8 71 1				
	transparency/opacity	transparent	transparent/opaque	only opaque		
compositionality	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.	major, idiosyncratic discrepancy between lexical meaning of component and WF		
com	binatorial ılarity / anomaly	formal/collocational regularity	metaphor) in components formal and/or collocational irregularity	semantics major combinatorial idiosyncrasy, anomaly		

 Table 9: Axes and degrees of idiomaticity

The findings of the study summarised above indicate that it is indeed possible to study singleword 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. References (selected):

- Bauer, L. (1983). English word-formation. Cambridge: Cambridge University Press.
- 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.
- 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.
- Biber, D. et al. (1999) Longman grammar of spoken and written English. Edinburgh: Pearson Education Ltd.
- 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. (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.
- 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). Frazeologie 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.

- Giegerich, H. J. (2004). Compound or phrase? English noun-plus-noun constructions and the stress criterion. English Language & Linguistics 8 (1), 1-24.
- 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.
- 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 perspectivem*. Amsterdam-Philadelphia: John Benjamins, 27-50.
- Haspelmath, M. (2002). Understanding Morphology. London: Hodder Arnold.
- Haspelmath, M. & Sims, A. (2010). Understanding Morphology. London: Hodder Education.
- 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.
- 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.
- Lakoff, G., & Johnson, M. (1980). *Metaphors We Live By*. Chicago: University of Chicago Press.
- Lakoff, G., & Johnson, M. (1980). *Metaphors We Live By*. Chicago: University of Chicago Press.
- Mel'čuk, I. A. (1976). *Das Wort* [Die Konversion als morphologisches Mittel]. München: W. Fink.
- 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.
- Plag, I. (2003) Word-formation in English. Cambridge: Cambridge University Press.
- 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.

Sources

- The British National Corpus (BNC) Hoffmann, S., & Evert, S. (1996). BNCweb (CQPedition). 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).

Publishing activities

Vašků, K. (2016). Ivan Poldauf - vzpomínka na slavnou dobu. *Časopis pro Moderní Filologii*, 98(2/2016), 300–302. Retrieved from http://casopispromodernifilologii.ff.cuni.cz/wp-content/uploads/sites/9/2016/09/katerina_vasku_300-302-1.pdf

Vašků, K. (2013). Alexander Onysko, Sascha Michel (eds): Cognitive Perspectives on Word Formation. *Časopis pro Moderní Filologii*, *95*(1/2013), 76–78.

Vašků, K. (2011a). Anglické vytýkací konstrukce a jejich české protějšky. In *Korpusová lingvistika Praha 2011 – 1 InterCorp*. (pp. 296–307). Praha: Nakladatelství Lidové noviny.

Vašků, K. (2011b). Pražská přednáška Dirka Geeraertse. *Studie z Aplikované Lingvistiky / Studies in Applied Linguistics*, 2011(2), 130–131. Retrieved from http://sali.ff.cuni.cz/?q=node/199#dirk

Vašků, K. (2011c). Vytýkací konstrukce v angličtině a v norštině. Kontrastivní studie. *Časopis pro Moderní Filologii*, 2011(2), 95–103.

Translation:

Klégr, A, a Vašků, K – Haspelmath, M., Sims. A.D. (2015) O čem je morfologie. Praha:

Karolinum., UK v Praze, s. 376, ISBN 978-80-246-2504-1