



FACULTY OF ARTS
Charles University

Otcí zakladatelé a matky samoživitelky.
**Frazeologické okolí vybraných rodových dvojic
v češtině a angličtině**

DENISA ŠEBESTOVÁ

ÚSTAV ANGLICKÉHO JAZYKA A DIDAKTIKY, FF UK

DENISA.SEBESTOVA@FF.CUNI.CZ

východiska: frazeologie

frekvenčně založená frazeologie:

- jazyk do velké míry spoléhá na opakující se zažité kombinace slov (např. Ebeling & Hasselgård 2015: 207)
- “composed of multi-word prefabricated expressions” (Biber et al. 2004: 372)
- idiom principle (Sinclair 1991: 110)
 - “the principle of idiom is that a language user has available to him a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analysable into segments.”

n-gramy kontrastivně

jazyky v kontrastu = ❤

+ n-gramy: velký potenciál (Granger 2014)

ale problematické – typologické odlišnosti (.....)

čeština X angličtina – zatím málo (Čermáková a Chlumská 2017)

typologické rozdíly: morfologie

CZ	EN
<i>za chvíliku</i>	<i>a little while later</i>
<i>chvíliku</i>	<i>for a little while</i>

typologické rozdíly: slovosled

CZ	EN
<p><i>Jan si dal čaj s mlékem</i> <i>Čaj s mlékem si dal Jan</i> <i>Čaj si dal Jan s mlékem</i> <i>Jan si čaj dal s mlékem</i></p> <p style="text-align: center;">...</p>	<p><i>Jan had tea with milk</i></p>

n-gramy a typologické odlišnosti

co s tím:

- lemmatizace (Čermáková & Chlumská 2017)
- skip-gramy (umožnit variabilní pořadí gramů) (Cheng et al. 2006)
- Engrammer! (Milička 2019)
 - <http://www.milicka.cz/en/engrammer/>

I: fraz. vyjádření času v dětské literatuře

3-4-gramy obsahující vybraná temporální jádra (autosémantika s jednoznačně časovou referencí, z top 500 lemmat)

vesměs odpovídající překladové protějšky č, a: *time – doba, chvíle – moment...*

dětská literatura – specifické zobrazení času? – Thompson & Sealey 2007: n-grams
“denoting times and places where narrative events unfold” (e.g. *for the first time, out of the window*)“

jazyk	korpus	subkorpus z	tokenů	textů
angličtina	BNC-jun	BNC (1994)	2 046 755	76 (úryvky)
čeština	SYN-7-jun	SYN-7 (2010)	2 821 044	59 (celé knihy)

frazeologie TIME

TIME	freq w/ TIME	total freq
for the first *	125	144
at the same *	93	110
for a long *	101	130
long * ago	24	26
* to time	19	20
a long * ago	21	21
from * to time	19	19
from time to *	19	19
for the second *	13	15
long * ago .	13	14



for the first *

time	125
emergence	1
skep	1
rung	1
miracle	1
act	1
couple (of weeks)	1
sign	1
few (years)	2
six (months)	1
hour	1
week	1
word	1
children	1
three (years)	1
night	1
two (days)	1
and (last time)	1
, (serious time)	1

II: rodové dvojice substantiv

potenciál odrážet kulturní specifika chápání a zobrazování genderů (Čermáková & Mahlberg 2019)

jazyk	korpus	mil. tokenů
angličtina	BNC (1994)	100
čeština	SYN 2015	100

rodové dvojice substantiv

CZ	EN
otec – matka	father – mother
syn – dcera	son – daughter
král – královna	king – queen
manžel – manželka	husband – wife

otec	matka	father	mother
dědit z otce na syna	dědit z matky na dceru	his father's died/death/funeral <i>(HER – mnohem níž)</i>	toddler/baby/womb , expectant
otec zakladatel	matka samoživitelka, azyl pro m.	founding father	foster, surrogate mother's voice mother and wife
	<i>verba dicendi</i>		queen mother, mother superior
	<i>idiom.</i> krkavčí matka		mother tongue

syn	dcera	son	daughter
měli syna a dceru měli dceru a syna	vztah matky a dcery, dcera a zet', dcera a vnučka	son and NUM daughters	son and NUM daughters
prvorzený kupcův předávat dědit	narodila se jim, vychovávali	<i>charakterizace:</i> eldest prodigal (marnotratný)	<i>přivlastnění.</i> Mr X's daughter daughter of Mr X
marnotratný, ztracený, svatý (<i>Bible</i>)		s. of a bitch	

král	královna	king	queen
<i>geogr. různorodé:</i> francouzský k. Ludvík uherský k. Matyáš španělský k. Filip český k. Přemysl Václav IV.	Alžběta	<i>více králů než královen</i> <i>geogr. diverzita</i> <i>English +</i> King Fahd Hussein of Jordan Juan Carlos	her Majesty the Q. Elizabeth Mary Q. of Scots Victoria the Queen's speech
ostrovy k. Jiřího náměstí """	země * maud letohrádek k. Anny	King's Cross King's College King's Road	Q.'s bench (division) Q's University Belfast award aircraft flight Hall The Queen Victoria
k. Lear k. Artuš, k. Herodes	k. ze Sáby		Fairy Queen

manžel	manželka	husband	wife
<i>generické užití v právní term. společné jmění manželů</i>	pojmout za m. ženatý, s manželkou X m. a dcera/2 děti	husband and wife	<i>no name</i> Mr X and his wife accompanied by his w. farmer's wife
její zesnulý její bývalý podváděný její budoucí	jeho manželka X	her late husband, her husband died, she killed her h., death of her h.	death of his wife
	m. nebo přítelkyně m. a milenka m. a matka	her first/second husband	his first/second wife

she killed her husband – výš na seznamu; s ničím nealternuje

ale *he killed his wife* 2x častější (44:25); *he killed his X = 1) wife, 2) father*

Right to left DefaultFont size 14 Show hints

wife

kill

 Words Lemmata
 Ordered Disordered
2 3 4 5 6 7
 Minimal number of texts 2

Risk ratio CI (risk of n-grams) 99.9 %

 Collapse similar results

N-gram Metric Ratio Txt

kill he * 175,096 44 / 966 31

N-gram	Metric	Ratio	Txt
kill he *	175,096	44 / 966	31
he * (80,9303	35 / 1537	35
he *,	43,5679	28 / 2123	23
he * .	39,8688	651 / 88160	406
he * ,'	36,0968	31 / 2921	29
he *	35,4864	7175 / 1197222	1452
he * ?	35,4015	44 / 4647	43
he * .'	14,8224	86 / 24877	70
. he *	13,6149	410 / 156493	278
' he *	5,37849	37 / 24443	36
. ' he *	5,32797	28 / 17221	28

Absolute frequency	
25.0 %	
Word	Metr Ratio
-	274 274 / 14087337
,	117 117 / 14468712
?	55 55 / 1147953
!	50 50 / 419487
wife	44 44 / 54984
and	37 37 / 7835997
in	18 18 / 5805744
father	17 17 / 71997
,	14 14 / 549609
if	10 10 / 757917
parent	9 9 / 58944
brother	9 9 / 35766
'	8 8 / 1448622
instantly	8 8 / 4560
with	8 8 / 1972899
son	7 7 / 48036
friend	7 7 / 92490
;	7 7 / 558561
mother	7 7 / 82821
family	6 6 / 125490
when	6 6 / 627426

závěry

Engrammer užitečný v porovnávání typologicky odlišných jazyků

- možná lemmatizace
- sloučení různých délek (problém nekorespondence délky ekvivalentních n-gramů)

odhaluje varianty alternující v 1 pozici > konstrukční pohled

frazeologické vzorce odrážejí rozdíly typologické, ale také (zajímavější) kulturní

díky!

denisa.sebestova@ff.cuni.cz

literatura

- Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad and Edward Finegan. 1999. *Longman Grammar of Spoken and Written English*. London: Longman.
- Cheng, W., Greaves, C., Warren, M. 2006. From n-gram to skipgram to concgram. *International Journal of Corpus Linguistics* 11(4), 411–433.
- Čermáková, A., Chlumská, L. 2017. Expressing PLACE in children's literature. Testing the limits of the n-gram method in contrastive linguistics. In: Dirdal, H., Egan, T. (eds) *Cross-linguistic Correspondences: From Lexis to Genre*, pp. 75–95. Benjamins, Amsterdam.
- Čermáková, Anna, and Michaela Mahlberg. 2019. Children's literature: Exploring the gendered social structure of fictional worlds. Conference paper. Corpus Linguistics 2019. Cardiff University, 26 July 2019.
- Ebeling, S.O. and Hasselgård, H. 2015. 'Learner corpora and phraseology.' In S. Granger, G. Gilquin and F. Meunier (eds.) *The Cambridge Handbook of Learner Corpus Research*. Cambridge: Cambridge University Press, 207-230.
- Granger, Sylviane. 2014. "A lexical bundle approach to comparing languages: Stems in English and French." *Languages in Contrast* 14:1, 58–72.
- Sinclair, John. 1991. *Corpus, Concordance, Collocation*. Oxford: OUP.
- Thompson, Paul & Alison Sealey. 2007. "Through children's eyes? Corpus evidence of the features of children's literature." *International Journal of Corpus Linguistics* 12:1, 1–23.

data

The British National Corpus, version 3 (BNC XML Edition). 2007. Distributed by Bodleian Libraries, University of Oxford, on behalf of the BNC Consortium. URL:
<http://www.natcorp.ox.ac.uk/>.

SYN2015. Representative Corpus of Contemporary Written Czech. Institute of the Czech National Corpus FF UK, Prague 2015. URL: <http://www.korpus.cz>.

Milička, J. 2019. *Engrammer*. <http://www.milicka.cz/en/engrammer/>