Probing sound and meaning with corpus data in Czech

Masako U. Fidler

Brown University

masako_fidler@brown.edu

Outline

1. Some relevant information about Czech
   a. Phonological inventory - orthography
   b. Morphological properties
   c. Position of onomatopoeic expressions (OpEs)
2. Definition of OpEs
3. Some samples of Czech OpEs
4. A divide between sound/motion-iconic expressions and the rest of the language
5. Research questions (and where they might lead to)
6. OpEs vs. grammaticalized forms of OpEs in texts (colligation, collocation patterns)
7. Discourse functions (collocates, keywords)
8. Phonosemantic relationships in OpEs and word formation (collocates)
9. Offshoot: selection process exists on every level: from OpE to OpE-based grammatical forms, OpEs and discourse function → Rethinking de Saussure's view

Some overview of Czech language

Czech phonological inventory: vowels

		front	central	back
high	short	/i/		/u/
	long	/iː/		/uː/
mid	short	/ɛ/		/o/
	long	/ɛː/		(/oː/)
low	short		/a/	
	long		/aː/	

Table 1: Czech vowel phonemes

Czech phonological inventory: consonants

			place of articulation							
			labial		alveo-dental		palatal		velo-glottal	
			bi-labial	labio-dental	apico-dental	apico-alveolar	apico-alveo-palatal	lamino-palatal	dorso-velar	glottal
obstruents	stop	-voiced	/p/		/t/			/c/	/k/	
		-voiced	/b/		/d/			/ɟ/	(/g/)	
	fricative	-voiced		/f/		/s/	/ʃ/		/x/	
		+voiced		/v/		/z/	/ʒ/[11]			/h/
	affricate	(+strident)				/ts/	/tʃ/			
	trill	+strident				/r̝/				
sonorants	trill	-strident				/r/				
	lateral					/l/				
	nasal		/m/		/n/			/ɲ/		
	glide							/j/		

Table 2: Czech consonant phonemes

Orthography

- Orthography largely reflects sound.
- Short: i/y, u, e, o, a
- Long: í/ý, ú/ů, é, ó, á
- Bi-labial and labio-dental: p, b, f, v, m
- Apico-dental: t, d, n
- Apico-alveolar: s, z, c, ř, r, l
- Apico-alveo-palatal: š, ž, č,
- Lamino-palatal: ť, ď, ň, j
- Dorso-velar: k, g, ch
- Glottal: h

Morphological properties (a rough outline)
Nominal inflection

- 7 grammatical cases (sg. and pl.) (standard grammatical presentation)

- Modifiers agree with the head noun in grammatical gender (masc, fem, neut), animacy (in masc), and grammatical number

Animate: *bratr* 'brother'

	sg	pl
N	*bratr*	*bratř***i**
G	*bratr***a**	*bratr***ů**
D	*bratr***ovi/-u**	*bratr***ům**
A	*bratr***a**	*bratr***y**
V	*bratř***e**	
L	*bratr***ovi/-u**	*bratr***ech**
I	*bratr***em**	*bratr***y**

Janda and Townsend (22)
http://www.seelrc.org:8080/grammar/pdf/stand_alone_czech.pdf

Morphological properties (a rough outline): Verbs

- Tense: non-past vs. past
- Agreement with 1-3 Person (& number) in the non-past
- Agreement with 1-3 person, number, & gender in the past tense
- Mood (imperative, conditional (present & (past))
- Verbal aspect: perfective vs. imperfective
- Verbal adverbs, verbal adjectives

Být 'to be' (Janda and Townsend 33)

	Present	Past	Future	Conditional
1sg	jsem	byl/-a/-o jsem	budu	byl/-a/-o bych
2sg	jsi	byl/-a/-o jsi/s	budeš	byl/-a/-o bys
3sg	je	byl/-a/-o	bude	byl/-a/-o by
1pl	jsme	byli/-y/-a jsme	budeme	byli/-y/-a bychom
2pl	jste	byl/-a/-i/-y/-a jste	budete	byl/-a/-i/-y/-a byste
3pl	jsou	byli/-y/-a	budou	byli/-y/-a by

Definition: Czech onomatopoeic expressions (OpEs)

- Closest to the description of the imitative sound symbolism and synesthetic sound symbolism in Hinton et al. (1994:2-3)
- Non-idiosyncratic (= conventionalized) expressions
- Cover sound, motion, and emotion (but not expletives based on non-sound/motion related expressions, e.g. *ježíšimarijá, hergot*)
- Uninflected (Discussions will later expand to derived OpE-*based* forms)
- Understudied in Slavic ("naiveté" (Kořínek 1934))

Czech OpEs expressing sound: samples

- Source
  - made by a physical apparatus of an animate entity
    - e.g. *bé* 'sheep's blurting, a loud cry of a child'
  - made by an animate entity that manipulates an inanimate entity
    - e.g. *brnk* 'sound made when a finger plucks a string'),
  - made directly by an inanimate entity
    - e.g. *cvak* (a short clicking sound of some hard object against some entity).

  - Size

    - Bum (human body falling) > pum (army guns)
- Properties of sound: e.g. loudness, speed, frequency
- Overlapping with motion (e.g. interaction with landing site)

Czech OpEs expressing motion

- type of moving entity (e.g. liquid *káp*)
- path (e.g. direct, roundabout itinerary, *houpy*),
- speed (e.g. fast *šup*),
- type of contact (e.g. the moving object shattered on contact *křách*),
- landing site (e.g. a liquid surface *žbluňk*).
- May simultaneously represent sound:
  - e.g. *hrk* 'a sudden movement or interruption and the sound accompanying the motion,' *kec* 'liquid dropping on the place (where it does not supposed to) or a fast and clumsy fall.'

Czech OpEs expressing emotion

The speaker's emotional state, e.g.

- reaction to pain or astonishment (e.g. *jé*) and joy (e.g. *juch*).
- can also simultaneously represent sound: e.g. *che* 'expression of contempt and the sound of laughter).

A divide between sound/motion-iconic expressions and the rest of the languages

- Ideophones: "a totally different kind of linguistic animal" (Diffloth 1976)

- Ideophones: "an entirely different realm from the familiar nouns, verbs, and particles, and from bound morphemes" (Alpher 1994).

- sound-iconic expressions are primitive/naïve (Kořínek 1934 and Fónagy 1999).

Research questions

- What's OpEs own space in language?
  - What are the differences between OpEs and OpE-based parts of speech (OpE-based verbs)?
  - What are the functions of OpEs in text?

OpEs as an integral part of language

Offshoots: question about what is arbitrary if OpEs have their own functions

OpEs vs. OpE-based verbs colligation (the instrumental case)

Instrumental

ranked 2 out of 7 cases for OpE-based *cvaknout* (cf. 6 out of 7 for OpE *cvak*

→ *Cause of* the sound

Richie fell out of [her] bed and cvak-ed on the switch [lit. clicked with the switch] on her stereo.

(rather than *what the sound was like*:

Then I leaned forward and hit it [the button] with a head.

  Cvak!

The door locks unlocked.)

Collocates for OpEs vs. OpE-based verbs
Nouns referring to sound source among 20 top collocates*

OpE cvak

4 collocates

- trigger, (camera) shutter (ranked 6)
- heel (8)
- lock (12)
- door (15)

Verb cvaknout

- 15 collocates
- switch (ranked 1)
- ratchet (2)
- firing pin (3)
- trigger, (camera) shutter (4)
- cigarette lighter (5)
- pen (6)
- ball-point pen (9)
- door handle (10)
- lock (12)
- beak (13)
- light switch (15)
- heel (16)
- scissors (17)
- camera (18)
- tooth (20)

*Cvrček 2013

collocates
cvak (OpE) vs. cvaknout (continued)

cvak

Adverbs

suddenly (14)

Verbs

click^{pf} (3)

hang up^{pf} (7)

do/make^{pf} (17)

Punctuation

! (16)

cvaknout

Adverbs

in vain (8),

drily (11),

quietly (19)

Verbs

-

Punctuation

-

OpE vs. OpEs

- Surprisal component in discourse (!, suddenly)
- Sound itself: "do/make^{pf}"

(a)   He reached out for the door.
   Cvak!
   The voices disappeared. As though he cut off the music.

(b) Without raising his head, he observes how a police officer lights up a cigarette with a Dupont lighter, cvak, and returns it to [his] pocket.

(c) [I] threw it [a piece of tar] on the chimney. It fell on the metal roof, did cvak and flew away into the dark.

- What is responsible for an action <nouns indicating the sound source

- A few hours later the door opened, the switch *cvak-ed* [turned on], and lights were turned on in the bathroom.

- And the result? When the lights went off and David's camera shutter *cvak-ed* [the last picture was taken] for the last time, it was clear that the [people's] favorite Müller consolidated his position at Lenka's hit-parade.

- They *cvak-ed* [validated] the tickets and sat on the last empty seats.

Discourse functions of OpEs

- Discourse boundary marking
- "Cliff-hanger" function
- Reporting that the situation has impact ("dummy objects") (Fidler 2015)
- Contributing to a multi-layered discourse structure (Fidler 2015)

Keywords (KWs)

Word forms obtained statistically by contrasting word frequencies of a text with word frequencies of a Reference Corpus

- Significance measured by chi-square, log-likelihood tests
- Ranking of KWs by Difference Index (DIN, Fidler and Cvrček 2015) = measures prominence of KWs

- word forms that occur in a text more frequently than expected by chance alone
- often closely connected to the overarching themes and genre of a text or set of texts. (Scott and Tribble 2006, Baker and Ellece 2011).
- Highly ranked KWs are expected to belong to the most striking parts of discourse

Prominence:
A sample from Karel Čapek's Water Sprite's Tale

Croaking sound of frogs: *kvak, kva, kvá*

- KW ranking

- Co-occurrence with finite verbs

Part of "complicating actions" (Labov)), leading to resolution (the "cliff-hanger" section) - Discourse marking

The Water Sprite's Tale (Karel Čapek) OpEs representing croaking sounds (out of 99 KWs)

ranking	keywords	Significance (LL)	DIN (prominence)
6	kvak	111.804	99.9953
7	kvá	36.519	99.994
12	kva	98.456	99.9819

-- OpEs are very prominent in the text

https://kwords.korpus.cz/
LL, min. 2, sig level .5 all significant KWs, Ref Corpus SYN2015

Place of action in text
and OpEs (Water Sprite's Tale)

Storyline
Beginning   end

	Tvar	ARF	Distribuce
6	kvak	1.917	
7	kvá	1.0041	
12	kva	1.0173	
28	naříkal	1.3065	
67	přinesl	1.2835	
72	podíval	1.1947	
80	udělal	1.2761	
84	šel	1.1931	

OpEs

Finite action verb forms
(excluding "to be" and verbs
of speech and asking)

Placement of OpEs
& co-occurrence with keywords

kvak
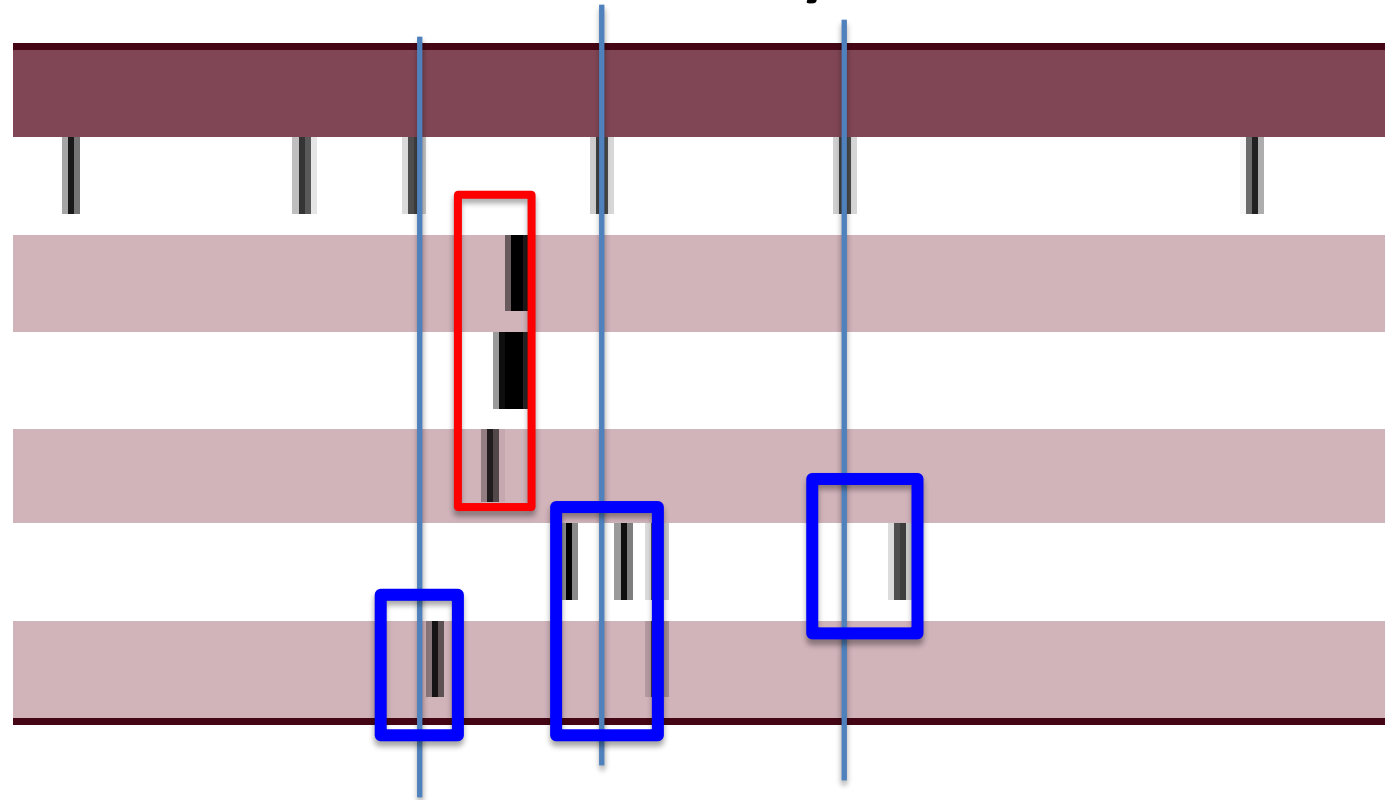kvá
kva
lamentedimpf
(into) air (motion)
(he) didpf

In more detail: discourse function

- "It was like this. Kvakvakvokoax fell in love. He saw Princess Kuakuakunka, and his heart was aflame, kvak. Kuakuakunka was beautiful. […] such a beauty."

- "[….] Kuakuakunka was beautiful and proud. She just puffed up and said kvak. Kvakvakvokoax was mad [with desire]. If you marry me, he said to her, I will bring and give you what you wish for. And so she said: Then bring me the blue of the sky, kvak."

- "What did he do then??" asked Zelinka. "Well, what could he have done? He sat under the water and lamented: Kva kva kva kvá, kva kva kva kvá. And then he wanted to take his own life. Therefore he jumped from the water into the air, so he would drown in it, kvak."

Another sample: *Stingy Bárka* (Jan Werich) bác (loud hitting sound) (ranked 1)

He caught the thief with a shovel in hand, how [dare] he tries to rob the needy! you bugger! You villain! And Kubát took the frail and BÁC, BÁC, BÁC! The thief fell. "Don't even move!" Of course, he didn't move. That is, she didn't move. Kubát recognized, only when he struck a match that it was not a male thief, but a female thief, and it was Barka from the Parish.

Bác as a discourse boundary marker

- They were in a good mood, it was just hahaha and all of a sudden *bác* and they were roaring: "Now you will speak!"

- Once she was on her way home and *prásk*, he was there.

Functions of OpEs

- As opposed to grammaticalized (inflected) OpE-based form
  - OpEs represent what the sound-motion is like
    (vs. what happened, what is responsible for the action)
- Discourse functions:
  - Marking prominent parts of text
  - Marking discourse chunks
    (can also indicate impact, create multilayered text)
- OpEs in Czech have their own functions in language (not random, disorganized forms)

OpE and OpE-based verbs
OpEs used as discourse markers

- Clicking sound of a lock, a camera shutter, clicking heels, turning on a switch (different sounds)

-- OpE cvak is associated with different sounds by way of shared *selected* properties.

- Turning on light, photographing, validating a ticket (inherently different sounds and different actions)

- -- *selected aspects* of the sounds lead to association with one verb: cvaknout

Arbitrary decision at every step

- OpEs and sound-motion-emotion, OpE-based forms and meaning, discourse functions of OpEs all involve some selection process
  - OpEs resemble direct quotation, which "demonstrates" rather than describes (Clark and Gerrig 1990) by highlighting some specific (and limited) aspects of the original utterance;
  - Process of grammaticalization: traces of lexical meaning remain when a lexical item becomes grammaticalized (Hopper 1991).
  - a pattern of evolution akin to "a termite's nest" without a "master plan" as linguistic forms are repeatedly used in contexts. (Bybee 2001)

Offshoots of this study

- Saussure's arbitrariness of sign might be a result of multiple selection process
- The relationship between sound and meaning might be deeply embedded in language
- Recent findings by Blasi, D E. et al. 2016 point in this direction.

References

Alpher, B. 1994. Yir-Yoront ideophones. Sound symbolism, ed. Leanne Hinton; Johanna Nichols;   and John Ohala, 161-77. Cambridge, UK: Cambridge UP.

Baker, P. and Ellece, S. 2011. *Key Terms in discourse analysis*. London: Continuum.

Blasi, D E. et al. 2016. Sound–meaning association biases evidenced across thousands of languages. Proceedings of the National Academy of Sciences of the United States of America. http://www.pnas.org/content/113/39/10818.full.pdf

Bybee, J. 2001. Phonology and Language Use. Cambridge.

Clark, H. H. and R. J. Gerrig. 1990. Quotation as demonstration. Language 66(4): 764–805.

Cvrček, V. 2013. *Kvantitativní analýza kontextu*. Prague: Lidové noviny.

Diffloth, G. 1976. Expressives in Semai. Austroasiatic studies, ed. by Philip N. Jenner; Laurence. C. Thompson; and Stanley Starosta. Part 1, Oceanic linguistics, Special publication 13, 249–264. Honolulu.

Fidler, M. Onomatopoeia in Czech: A conceptualization of sound and its connections to grammar and discourse. Bloomigton, IN: Slavica.

Fidler, M. and Cvrček. 2015. A fata-driven analysis of reader viewpoints: Reconstructing the historical reader using keyword analysis". *Journal of Slavic Linguistics* 23(2): 197–239.

Fónagy, I. 1999. Why iconicity? Form miming meaning, ed. Max Nänny and Olga Fischer, 3–35. Amsterdam.

Hinton, Leanne; Nichols, Johanna; and Ohala, John J. Eds. 1994. Sound symbolism. Cambridge,

UK: Cambridget UP.

_____. 1994. Introduction: Sound symbolic processes. Hinton et al. 1-14.

Hopper, Paul. 1991. On some principles of grammaticization". Approaches to Grammaticalization, Vol. I., eds. Elizabeth Closs Traugott and Bernd Heine, 17–36. Amsterdam.

Janda, L. A. and Ch. Townsend. Czech. http://www.seelrc.org:8080/grammar/pdf/stand_alone_czech.pdf

Kořínek, Josef Miroslav. 1934. Studie z oblasti onomatopoje. Prague.

Labov, W. *Language in the Inner City.* Philadelphia: Univ. of Pennsylvania Press.

Scott, M. and C. Tribble. 2006. *Textual patterns: Keyword and corpus analysis in language education*. Amsterdam: Benjamins.

Sources

Čapek, K. 1991. Devatero pohádek. Prague: Albatros.

Werich, J. 1960. Finfárum. Prague: Československý spisovatel.

Kwords (available at https://kwords.korpus.cz/)

SYN2015: Křen, Michal and Václav Cvrček et al. (2015) *SYN2015: Representative Corpus of Contemporary Written Czech*. Czech National Corpus, Faculty of Arts, Charles University, Prague. Available at: http://www.korpus.cz.