Probing sound and meaning with corpus data in Czech: A pilot study

NINJAL International Symposium 2016
Mimetics in Japanese and Other Languages of the World (Dec 17-18, 2016)

Masako U. Fidler
Brown University
masako_fidler@brown.edu
Outline

1. A short overview of Czech
2. Definition of Czech onomatopoeic expressions (OpEs)
3. Some samples of Czech OpEs
4. Onomatopoeia/Ideophones/mimetics and the “rest” of the language and the situation in Czech
5. Research question (and where they might lead to)
6. OpEs vs. grammaticalized forms of OpEs in Czech texts (colligation, collocation patterns)
7. Discourse functions of OpEs (collocates, keywords) in Czech
8. Interim observations
Outline

1. A short overview of Czech
2. Definition of Czech onomatopoeic expressions (OpEs)
3. Some samples of Czech OpEs
4. Onomatopoeia/Ideophones/mimetics and the “rest” of the language and the situation in Czech
5. Research question (and where they might lead to)
6. OpEs vs. grammaticalized forms of OpEs in Czech texts (colligation, collocation patterns)
7. Discourse functions of OpEs (collocates, keywords) in Czech
8. Interim observations
A short overview of Czech

https://www.online.muni.cz/tema/2972-nareci-nenavratne-mizi

Bohemian dialects

Moravian dialects
Czech phonological inventory: vowels

<table>
<thead>
<tr>
<th></th>
<th>front</th>
<th>central</th>
<th>back</th>
</tr>
</thead>
<tbody>
<tr>
<td>high</td>
<td>short</td>
<td>/i/</td>
<td>/u/</td>
</tr>
<tr>
<td></td>
<td>long</td>
<td>/iː/</td>
<td>/uː/</td>
</tr>
<tr>
<td>mid</td>
<td>short</td>
<td>/ɛ/</td>
<td>/o/</td>
</tr>
<tr>
<td></td>
<td>long</td>
<td>/ɛː/</td>
<td>(/oː/)</td>
</tr>
<tr>
<td>low</td>
<td>short</td>
<td></td>
<td>/a/</td>
</tr>
<tr>
<td></td>
<td>long</td>
<td></td>
<td>/aː/</td>
</tr>
</tbody>
</table>

Table 1: Czech vowel phonemes
## Czech phonological inventory: consonants

<table>
<thead>
<tr>
<th>manner of articulation</th>
<th>place of articulation</th>
<th>labial</th>
<th>alveo-dental</th>
<th>palatal</th>
<th>velo-glottal</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>bi-labial</td>
<td>labio- dental</td>
<td>apico- dental</td>
<td>apico- alveolar</td>
</tr>
<tr>
<td>obstruents</td>
<td></td>
<td>/p/</td>
<td>/t/</td>
<td>/c/</td>
<td>/k/</td>
</tr>
<tr>
<td></td>
<td></td>
<td>/b/</td>
<td>/d/</td>
<td>/j/</td>
<td>(/g/)</td>
</tr>
<tr>
<td>fricative</td>
<td></td>
<td>/f/</td>
<td>/s/</td>
<td>/ʃ/</td>
<td>/x/</td>
</tr>
<tr>
<td>+voiced</td>
<td></td>
<td>/v/</td>
<td>/z/</td>
<td>/ʒ/</td>
<td>/h/</td>
</tr>
<tr>
<td>+strident</td>
<td></td>
<td>/ts/</td>
<td>/tʃ/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>trill</td>
<td>-strident</td>
<td>/ɾ/</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>lateral</td>
<td></td>
<td>/l/</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>nasal</td>
<td></td>
<td>/m/</td>
<td>/n/</td>
<td>/ŋ/</td>
<td>/j/</td>
</tr>
<tr>
<td>glide</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Animate: bratr ‘brother’</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>sg</td>
<td>N</td>
<td>bratr</td>
</tr>
<tr>
<td>pl</td>
<td>G</td>
<td>bratra</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>bratrovi/-u</td>
</tr>
<tr>
<td></td>
<td>A</td>
<td>bratra</td>
</tr>
<tr>
<td></td>
<td>V</td>
<td>bratr ě</td>
</tr>
<tr>
<td></td>
<td>L</td>
<td>bratrovi/-u</td>
</tr>
<tr>
<td></td>
<td>I</td>
<td>bratrem</td>
</tr>
</tbody>
</table>

Janda and Townsend (22)

Morphological properties (a rough outline): Verbs

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

Být ‘to be’ (Janda and Townsend 33)

<table>
<thead>
<tr>
<th></th>
<th>Present</th>
<th>Past</th>
<th>Future</th>
<th>Conditional</th>
</tr>
</thead>
<tbody>
<tr>
<td>1sg</td>
<td>jsem</td>
<td>byl/-a/-o jsem</td>
<td>budu</td>
<td>byl/-a/-o bych</td>
</tr>
<tr>
<td>2sg</td>
<td>jsi</td>
<td>byl/-a/-o jsi/s</td>
<td>budeš</td>
<td>byl/-a/-o bys</td>
</tr>
<tr>
<td>3sg</td>
<td>je</td>
<td>byl/-a/-o</td>
<td>bude</td>
<td>byl/-a/-o by</td>
</tr>
<tr>
<td>1pl</td>
<td>jsme</td>
<td>byl/-y/-a jsme</td>
<td>budeme</td>
<td>byl/-y/-a bychom</td>
</tr>
<tr>
<td>2pl</td>
<td>jste</td>
<td>byl/-a/-i/-y/-a jste</td>
<td>budete</td>
<td>byl/-a/-i/-y/-a byste</td>
</tr>
<tr>
<td>3pl</td>
<td>jsou</td>
<td>byl/-y/-a</td>
<td>budou</td>
<td>byl/-y/-a by</td>
</tr>
</tbody>
</table>
Outline

1. A short overview of Czech
2. Definition of Czech onomatopoeic expressions (OpEs)
3. Some samples of Czech OpEs
4. Onomatopoeia/Ideophones/mimetics and the “rest” of the language and the situation in Czech
5. Research question (and where they might lead to)
6. OpEs vs. grammaticalized forms of OpEs in Czech texts (colligation, collocation patterns)
7. Discourse functions of OpEs (collocates, keywords) in Czech
8. Interim observations
Definition: Czech onomatopoetic expressions (OpEs)

- Closest to the description of the imitative sound symbolism and synesthetetic 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
- Understudied
Outline

1. A short overview of Czech
2. Definition of Czech onomatopoeic expressions (OpEs)
3. Some samples of Czech OpEs
4. Onomatopoeia/Ideophones/mimetics and the “rest” of the language and the situation in Czech
5. Research question (and where they might lead to)
6. OpEs vs. grammaticalized forms of OpEs in Czech texts (colligation, collocation patterns)
7. Discourse functions of OpEs (collocates, keywords) in Czech
8. Interim observations
Czech OpEs indicating 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. brn̩k ‘a sound made when a finger plucks a string’
  – made directly by an inanimate entity
    • e.g. cvak ‘a short clicking sound of a hard object in contact with another hard object’.

• Size
  • bum (hitting with/falling of a large entity, e.g. a human body) vs. pum (hitting with a bullet)
Czech OpEs indicating motion

- type of moving entity (e.g. liquid kář, small solid entity cup)
- Path and speed (e.g. roundabout itinerary, houp(y) (vs. hop), fast šup)
- type of contact (e.g. moving object shattered on contact křáč, bouncing object on contact pink)
- landing site (e.g. an entity hitting a liquid surface žbluňk, stepping on a hard surface dup).
- OpEs indicating motion can simultaneously represent sound:
  - e.g. hrk ‘a sudden movement or interruption and the sound poss. accompanying the motion,’ kec ‘liquid dropping on a place (where it is not supposed to) or a fast and clumsy fall.’
Czech OpEs expressing emotion

The speaker’s emotional state, e.g.

• reaction to pain (e.g. *jau*), astonishment (e.g. *jé*), joy (e.g. *juch*).

• can also simultaneously represent sound: e.g. *che* ‘expression of contempt and the sound of condescending jeer.’
Outline

1. A short overview of Czech
2. Definition of Czech onomatopoeic expressions (OpEs)
3. Some samples of Czech OpEs
4. Onomatopoeia/Ideophones/mimetics and the “rest” of the language and the situation in Czech
5. Research question (and where they might lead to)
6. OpEs vs. grammaticalized forms of OpEs in Czech texts (colligation, collocation patterns)
7. Discourse functions of OpEs (collocates, keywords) in Czech
8. Interim observations
A divide between ideophones/mimetic/onomatopoeic expressions and “the rest” of the language

- **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).
- **Primitive/naïve** (Kořínek 1934 and implicitly in Fónagy 1999).
- **The contrast between OpEs and the rest of the language:** absence of inflection
Outline

1. A short overview of Czech
2. Definition of Czech onomatopoeic expressions (OpEs)
3. Some samples of Czech OpEs
4. Onomatopoeia/Ideophones/mimetics and the “rest” of the language and the situation in Czech
5. Research question (and where they might lead to)
6. OpEs vs. grammaticalized forms of OpEs in Czech texts (colligation, collocation patterns)
7. Discourse functions of OpEs (collocates, keywords) in Czech
8. Interim observations
**Question**

What is the position of OpEs in Czech?

If OpEs are uninflected,

– Does it mean that OpEs are not part of the “system”, a disorganized group of word forms?

– What are the differences between OpEs and OpE-based parts of speech (OpE-based verbs)?

– What are the functions of OpEs in discourse?

Czech OpEs as an integral part of language

Would studying OpEs help us rethink the arbitrariness principle by Saussure?
Outline

1. A short overview of Czech
2. Definition of Czech onomatopoeic expressions (OpEs)
3. Some samples of Czech OpEs
4. Onomatopoeia/Ideophones/mimetics and the “rest” of the language and the situation in Czech
5. Research question (and where they might lead to)
6. Division of labor between OpEs vs. grammaticalized forms of OpEs in Czech texts (colligation, collocation patterns)
7. Discourse functions of OpEs (collocates, keywords) in Czech
8. Interim observations
A pilot study
OpEs vs. OpE-based verbs
(sample: cvak vs. cvaknout)
colligation

Instrumental
ranked 2 out of 7 cases for OpE-based cvaknout
vs. 6 out of 7 for OpE cvak

→ How (by means of what) something is done, what happens
‘Richie fell out of [her] bed and cvak-ed on the switch\textsuperscript{inst} [lit. clicked with the switch] on her stereo.’ = turned on on the switch

vs. what the sound is like
‘Then I leaned forward and hit it [the button] with a head\textsuperscript{instr}. 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)

*within the context of (-3, +3) Cvrček 2013
Collocates

cvak (OpE) vs. cvaknout (continued)

cvak

Verbs
-click\textsuperscript{pf} (cvaknout) (3)
-hang up\textsuperscript{pf} (7)
-do/make\textsuperscript{pf} (17) (cf. Hamano 1998, -suri)

Adverbs
-suddenly (14)

Punctuation
-! (16)

cvaknout

Verbs
-

Adverbs
-in vain (8),
-drily (11),
-quietly (19)

Punctuation
-
Qualitative examples

CVAK
• Marker for a surprisal moment in discourse (!, suddenly)
• Type of sound itself

CVAKNOUT
• What happens, how it happens

da) A few hours later the door opened, the switch *cvak-ed* [turned on], and there was light in the bathroom.

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

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.

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

f) He finally understood it. [lit. It finally *cvak-ed* for/on him]

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.
Outline

1. A short overview of Czech
2. Definition of Czech onomatopoeic expressions (OpEs)
3. Some samples of Czech OpEs
4. Onomatopoeia/Ideophones/mimetics and the “rest” of the language and the situation in Czech
5. Research question (and where they might lead to)
6. Division of labor between OpEs vs. grammaticalized forms of OpEs in Czech texts (colligation, collocation patterns)
7. Discourse functions of OpEs (collocates, keywords) in Czech
8. Interim observations
9. Offshoots: selection process exists on every level: from OpE to OpE-based grammatical forms, OpEs and discourse function \(\rightarrow\) Rethinking de Saussure’s view
Discourse functions of OpEs

1. Discourse boundary marking (surprisal moment in discourse – “cliffhanger moment”)

2. Reporting that the situation has impact (“dummy objects”) (Fidler 2014)

3. Contributing to a multi-layered discourse structure (Fidler 2014)
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 effect size - Difference Index (DIN, Fidler and Cvrček 2015) = measures prominence of KWs
  – https://kwords.korpus.cz/
• 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 the discourse
Prominence:
A sample from Karel Čapek’s Water Sprite’s Tale

Croaking sounds of frogs: kvak, kva, kvá

• KW ranking

• Co-occurrence with keyed finite verb forms (presumably part of “complicating actions” (Labov 1972))
The Water Sprite’s Tale (Karel Čapek)
OpEs representing croaking sounds
(out of 99 KWs)

---

<table>
<thead>
<tr>
<th>ranking</th>
<th>keywords</th>
<th>Significance (LL)</th>
<th>DIN (prominence)</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>kvak</td>
<td>111.804</td>
<td>99.9953</td>
</tr>
<tr>
<td>7</td>
<td>kvá</td>
<td>36.519</td>
<td>99.994</td>
</tr>
<tr>
<td>12</td>
<td>kva</td>
<td>98.456</td>
<td>99.9819</td>
</tr>
</tbody>
</table>

---

-- OpEs are very prominent in the text

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

OpEs

Finite action verb forms (excluding “to be” and verbs) of speech and asking)
The most critical moment in the last segment of the text

kvak
kvá
kva

lamented\textsuperscript{impf}

(into) air \textit{(motion)}

(he) did\textsuperscript{pf}
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:

*The Stingy Bárka* (Jan Werich) and OpE *bác* (a loud hitting sound)

*Bác* – (loud hitting sound): ranked 1 (DIN=100)

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 Bárka from the Parish.
OpE, used 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.
Outline

1. A short overview of Czech
2. Definition of Czech onomatopoeic expressions (OpEs)
3. Some samples of Czech OpEs
4. Onomatopoeia/Ideophones/mimetics and the “rest” of the language and the situation in Czech
5. Research question (and where they might lead to)
6. Division of labor between OpEs vs. grammaticalized forms of OpEs in Czech texts (colligation, collocation patterns)
7. Discourse functions of OpEs (collocates, keywords) in Czech
8. Interim observations
Interim observations

• As opposed to grammaticalized (inflected) OpE-based forms:
  – OpEs represent what the sound-motion-emotion is like
    (vs. what happens, by means of what it happens)

• OpEs have discourse functions:
  – Marking prominent parts of text
  – Marking discourse chunks
    (can also indicate impact, create layers in text)

• OpEs in Czech have their own functions in language
  (not random, disorganized forms that do not belong to
  the language because they don’t seem to belong to the
  inflectional system)
Returning to: *Cvak* vs. *cvaknout*

OpEs used as discourse markers

- Clicking sound of a lock, a camera shutter, clicking heels, turning on a switch
- These are different naturally occurring sounds
  -- OpE *cvak* is chosen to represent different sounds by way of shared *selected* properties;

OpEs resemble direct quotation, which “demonstrates” rather than describes (Clark and Gerrig 1990) by highlighting some specific (and limited) aspects of the original utterance.

- Turning on a light, photographing, validating a ticket, coming to understanding something
- Different actions, with sound or without sound
  -- also a level of selection: *cvaknout* is chosen to represent different actions (loosely connected with the clicking sound)
Outline

1. A short overview of Czech
2. Definition of Czech onomatopoeic expressions (OpEs)
3. Some samples of Czech OpEs
4. Onomatopoeia/Ideophones/mimetics and the “rest” of the language and the situation in Czech
5. Research question (and where they might lead to)
6. Division of labor between OpEs vs. grammaticalized forms of OpEs in Czech texts (colligation, collocation patterns)
7. Discourse functions of OpEs (collocates, keywords) in Czech
8. Interim observations
It is possible that a selection process is made at every step

- Matching of the selected aspects of sound segments of an OpE and naturally occurring sounds-motions-emotions
- Matching of OpE-based forms and events
- These processes all seem to involve a selection process (picking shared features and matching with one form)
- Such repeated selection processes are commensurate with the “termite’s nest” concept of the evolution of language
  - A pattern of language evolution akin to “a termite's nest” without a “master plan” as linguistic forms are repeatedly used in contexts. (Bybee 2001)
Sound and meaning may be deeply embedded

• Saussure’s arbitrariness of sound-meaning may be seen as a snap shot. It may simply reflect a result of multiply repeated selection processes
Instead, the relationship between sound and meaning might be in fact deeply embedded in language but we don’t see it if we take a static snapshot of the majority of linguistic forms.

• Basic relationships between sound and meaning - cf. Recent findings by Blasi, D E. et al. 2016 pointing in this direction: “Sound–meaning association biases evidenced across thousands of languages”
References


Sources


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

Thank you!