



ČESKÝ NÁRODNÍ
KORPUS

Korpusové přístupy k analýze diskurzu

Václav Cvrček
36. studentský workshop Žďárek
29. dubna 2016





Úvod



Úvodem

Interpretace textu

- ▶ jádro práce humanitního vědce
- ▶ každá interpretace je hodna pozornosti
- ▶ objektivní interpretace neexistuje (vždy záleží na úhlu pohledu)
 - ▶ snaha o minimalizaci předpojatosti (*researcher bias*)
 - ▶ empirické zakotvení (minimum intuice)
 - ▶ vycházející z textu (minimum vnětextových informací)





Prominentní jednotky



Interpretace a prominence

Kde začít s interpretací textu?

- ▶ co je na textu zvláštního/specifického?
- ▶ témata, motivy – vyjadřené slovy
- ▶ interakce mezi slovy/tématy
- ▶ funkce a význam slov/motivů

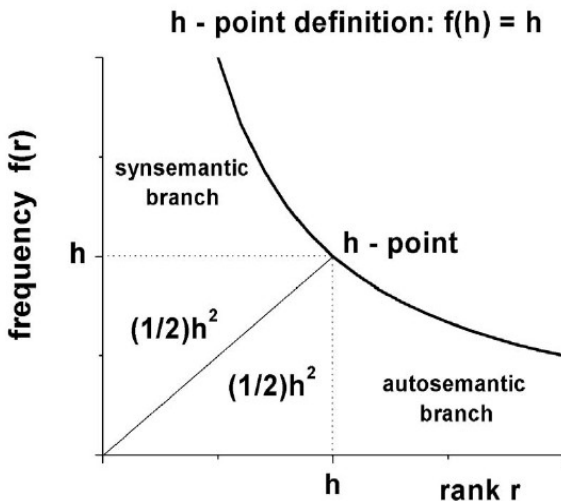
CADS = corpus-assisted discourse studies

Tématická koncentrace

- ▶ plnovýznamová slova s „abnormální“ frekvencí
- ▶ založeno na I. Zipfově zákonu ($\text{rank} \times \text{fq} = \text{konst.}$)
 - ▶ h -point: $\text{rank} = \text{frekvence}$
 - ▶ h -point – přibližně odděluje synsémantickou a autosémantickou složku distribuce
 - ▶ plnovýznamová slova nad h -bodem jsou TC
- ▶ © Popescu & Altmann 2006, David et al. 2013, Čech 2016



Tématická koncentrace



Tématická koncentrace – příklady

1984:

Winston, say, know, Party, face, word, O'Brien, seem, look, never, think, moment, always, hand, year, way, long, now, eye, day, possible, war...

SOTU:

year, job, work, America, new, people, american, know, need, help, country, business, time, world, economy, family, right, tax, Congress, nation...

Hobbit:

come, go, Bilbo, see, dwarf, time, long, make, think, great, good, know, far, still, goblin, find, way, look, little, light...



Tématická koncentrace – diskuse

Klady a zápory tématické koncentrace

- + objektivita – založeno na frekvenční distribuci
- + aplikace: porovnání textů na základě tématické kompaktnosti
- + nevyžaduje referenční korpus
- množina prominentních jednotek je invariantní
- „interpretace bez osoby interpreta“ – pomíjí úhel pohledu





Analýza klíčových slov



Analýza klíčových slov

Klíčová slova (KWs)

- ▶ terminologické homonymum (!)
- ▶ *A word-form which recurs within the text in question will be more likely to be key in it.* (M. Scott 2006)
- ▶ srovnání relativní frekvence v textu s referenčním korpusem
- ▶ referenční korpus představuje modelového příjemce
- ▶ testy signifikance: χ^2 test, log-likelihood (G) test, Fisher test
- ▶ určení míry prominence (effect size – DIN)
- ▶ klíčová slova jsou východiskem interpretace (nikoli cílem analýzy)



Analýza klíčových slov (KWA)

Romeo and Juliet vs. all Shakespeare plays (Scott et al. 2006)

AH	DEATHLY	MARRIED	SLAIN
ART	EARLY	MERCUTIO	THEE
BACK	FRIAR	MONTAGUE	THOU
BANISHED	JULIET	MONUMENT	THURSDAY
BENVOLIO	JULIET'S	NIGHT	THY
CAPULET	KINSMAN	NURSE	TORCH
CAPULETS	LADY	O	TYBALT
CAPULET'S	LAWRENCE	PARIS	TYBALT'S
CELL	LIGHT	POISON	VAULT
CHURCHYARD	LIPS	ROMEO	VERONA
COUNTY	LOVE	ROMEO'S	WATCH
DEAD	MANTUA	SHE	WILT

Algoritmus pro identifikaci KWs

Postup

- ▶ zjistit frekvenci všech slov v textu – nejfrekventovanější jsou *a*, *se na*, *v...*
- ▶ porovnat relativní frekvence jednotek v textu a v referenčním korpusu
- ▶ použít statistický test: χ^2 , log-likelihood nebo Fisherův exaktní test pro zjištění, zda je rozdíl signifikantní
- ▶ interpretovat prvních X nejvíc prominentních jednotek



Poznámka o signifikanci a relevanci

Gabrielatos, C. & Marchi, A. (2012): je rozdíl mezi statistickou signifikancí a relevancí (effect size)

Metrika k určování prominence/keyness

- ▶ **signifikance** – míra jistoty, že máme dostatek dat pro konstatování nenáhodného rozdílu
- ▶ **relevance** – významnost rozdílu pro interpretaci
- ▶ zásadní při interpretaci pouze prvních X klíčových slov:
 1. **identifikace** KWs – statistické testy
 2. **pořadí** KWs – odlišná metrika

Koeficient DIN

Variace na Sørensen–Diceův koeficient¹:

$$DIN = 100 \times \frac{RelFq(Target) - RelFq(Reference)}{RelFq(Target) + RelFq(Reference)}$$

- ▶ hodnoty DIN
 - ▶ -100 (= slovo se vyskytuje pouze v RefC)
 - ▶ 0 (= slovo se vyskytuje stejně často v textu a v RefC)
 - ▶ 100 (= slovo je pouze ve zkoumaném textu)
- ▶ DIN = rozdíl relativních frekvencí k jejich průměru ($\times 50$)
- ▶ hodnota 100 pro všechna slova, která nejsou v RefC (!)
- ▶ užitečné pro stanovování pořadí, nikoli pro identifikaci

¹cf. Hofland–Johansson (1982).

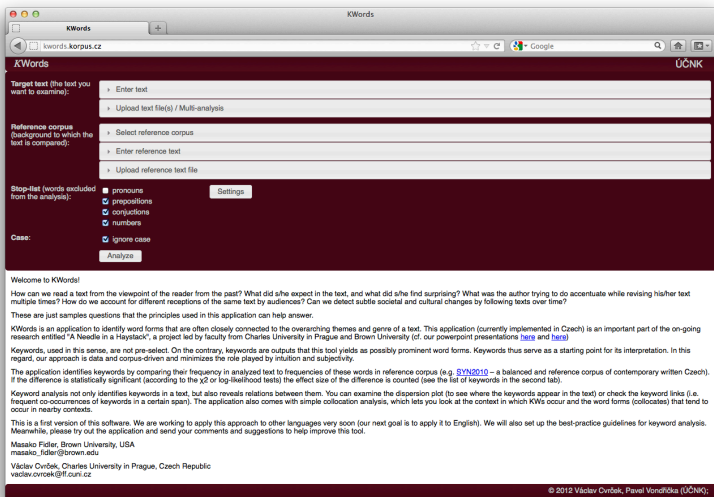


Nástroje pro KWA



KWords

<http://kwords.korpus.cz>



The screenshot shows a web browser window with the URL kwords.korpus.cz. The page title is "KWords" and the logo "ÚČNK" is in the top right corner. The interface is divided into several sections:

- Target text (the text you want to examine):** Includes a text input field with a "Enter text" button and an "Upload text file(s) / Multi-analysis" button.
- Reference corpus (background to which the text is compared):** Includes a "Select reference corpus" button, an "Enter reference text" input field with a "Enter reference text" button, and an "Upload reference text file" button.
- Stop-list (words excluded from the analysis):** A "Settings" button and a list of checkboxes: prepositions, conjunctions, numbers, and ignore case.
- Case:** A checkbox for ignore case and an "Analyze" button.

Below the form is a "Welcome to KWords!" section with introductory text:

How can we read a text from the viewpoint of the reader from the past? What did s/he expect in the text, and what did s/he find surprising? What was the author trying to do accentuate while revising his/her text multiple times? How do we account for different receptions of the same text by audiences? Can we detect subtle societal and cultural changes by following texts over time?

These are just samples questions that the principles used in this application can help answer.

KWords is an application to identify word forms that are often closely connected to the overarching themes and genre of a text. This application (currently implemented in Czech) is an important part of the on-going research entitled "A Needle in a Haystack", a project led by faculty from Charles University in Prague and Brown University (cf. our powerpoint presentations [here](#) and [here](#)).

Keywords, used in this sense, are not pre-select. On the contrary, keywords are outputs that this tool yields as possibly prominent word forms. Keywords thus serve as a starting point for its interpretation. In this regard, our approach is data and corpus-driven and minimizes the role played by intuition and subjectivity.

The application identifies keywords by comparing their frequency in analyzed text to frequencies of these words in reference corpus (e.g. [SYN2010](#) – a balanced and reference corpus of contemporary written Czech). If the difference is statistically significant (according to the χ^2 or log-likelihood tests) the effect size of the difference is counted (see the list of keywords in the second tab).

Keyword analysis not only identifies keywords in a text, but also reveals relations between them. You can examine the dispersion plot (to see where the keywords appear in the text) and check the keyword links (i.e. frequent co-occurrences of keywords in a certain span). The application also comes with simple collocation analysis, which lets you look at the context in which KWs occur and the word forms (collocates) that tend to occur in nearby contexts.

This is a first version of this software. We are working to apply this approach to other languages very soon (our next goal is to apply it to English). We will also set up the best-practice guidelines for keyword analysis. Meanwhile, please try out the application and send your comments and suggestions to help improve this tool.

Masako Fidler, Brown University, USA
masako_fidler@brown.edu

Václav Cvrček, Charles University in Prague, Czech Republic
vacek.cvrcek@ff.cuni.cz

© 2012 Václav Cvrček, Pavel Vondříčka (ÚČNK)

KWords: analyzovaný text

KWords ÚČNK

Target text (the text you want to examine):

▶ Enter text

▶ Upload text file(s) / Multi-analysis

Reference corpus (background to which the text is compared):

▶ Select reference corpus

▶ Enter reference text

▶ Upload reference text file

Stop-list (words excluded from the analysis):

pronouns

prepositions

conjunctions

numbers

ignore case

Settings

Case:

Analyze

Text Keywords Distribution Keyword links Concordance

Vážené soudružky, vážení soudruzi, **drazí** spoluobčané!

Vstoupili jsme do nového **roku** 1975. Dovolte mi, abych vás všechny na jeho prahu **jménem** ústředního výboru **Komunistické strany Československa**, ústředního výboru Národní fronty, **jménem** československé vlády i nemocného prezidenta republiky a **jménem** svým srdečně pozdravil a popřál vám v novém **roce** hodně zdraví, štěstí, **osobní** spokojenosti i mnoho nových pracovních úspěchů.

V životě **člověka** i v životě **společnosti** a **státu** je to příležitost k zamyšlení, co se nám v **minulém roce** zdařilo a co ne, co **nás** očekává a co si slibujeme od příštího **roku**.

Můžeme říci, že uplynulý **rok** 1974 byl ve **všech oblastech** pro Československo dobrým **rokem**. Důstojně se přidal k několika předcházejícím rokům v rozvoji **našeho hospodářství**, **životní** úrovně a **celé společnosti**. Dobře pracoval **náš** průmysl, stavebnictví, zemědělství, příznivé jsou **výsledky** v ostatních **oblastech**. Úspěšně jsme zakončili čtvrtý **rok** pátého pětiletého plánu.

Období, kterým jsme prošli od XIV. sjezu **Komunistické strany Československa**, je charakterizováno příznivým vnitropolitickým vývojem, upevňováním **našeho socialistického státu**, rozvojem **socialistické** demokracie, politickou angažovaností i pracovní aktivitou dělnické třídy, rolnictva, inteligence, žen, mládeže i prohlubováním morálně politické jednoty **pracujících lidí** i obou bratrských **národů** a **všech** národností.

KWords: seznam klíčových slov

2 typy prominentních jednotek: **klíčová slova** and **tematická koncentrace**

Text Klíčová slova Distribuce Keyword links Konkordance

Shrnutí

	Vstupní text	Referenční korpus: syn2010
Tokeny	1719	121666414
Typy	789	1706344

Tematická koncentrace

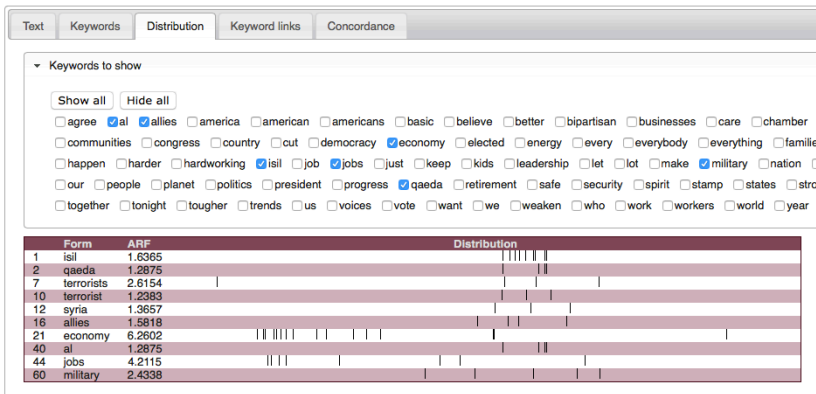
Pořadí	Tvar	FQ
7	lidu	15
10	roce	11

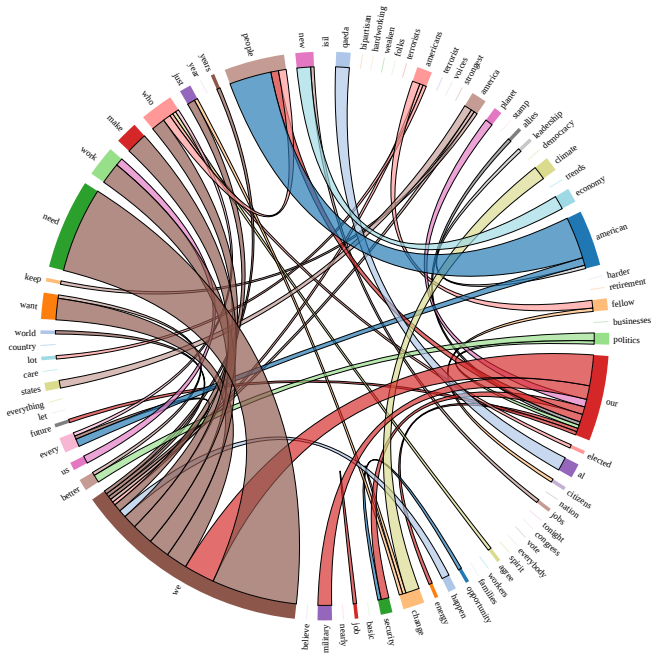
Klíčová slova

	Tvar	LL	DIN	Fq(text)	Fq(ref)	TC
1	drazí	49.773	99.8553	4	205	-
2	pracujícího	36.426	99.8315	3	179	-
3	svazem	42.521	99.6397	4	511	-
4	socialistického	53.132	99.6390	5	640	-
5	sovětským	42.475	99.6375	4	514	-
6	národně	31.787	99.6333	3	390	-
7	lidu	149.003	99.4889	15	2720	+
8	pracujících	28.647	99.3803	3	660	-
9	československého	27.194	99.2100	3	842	-
10	hospodářství	87.706	99.0847	10	3254	-
11	socialistické	25.765	98.9972	3	1070	-
12	zeměmi	25.743	98.9935	3	1074	-
13	surovin	25.222	98.9021	3	1172	-
14	komunistické	41.057	98.7889	5	2156	-
15	československa	32.726	98.7706	4	1751	-

Disperze klíčových slov

SOTU 2016: *terrorism* × *economy*






Multi-analýza

Pro srovnávání několika textů – časová řada (Novoroční projevy G. Husáka)

Text	Keywords	Distribution	Keyword links	Concordance	Comparison												
	Keywords	Match	first	second	third	fourth	fifth	sixth	seventh	eighth	ninth	tenth	eleventh	twelfth	thirteenth	fourteenth	fivteenth
lidu	14	1.990	1.985	1.981	1.991	1.983	1.983	1.986	1.985	1.987	1.988	1.977	1.984	1.950	1.970	-	-
roku	14	1.555	1.431	1.683	1.581	1.605	1.599	-	1.462	1.618	1.503	1.693	1.503	1.640	1.478	1.591	-
socialistické	14	1.980	1.988	1.990	1.989	1.991	-	1.990	1.988	1.993	1.989	1.986	1.987	1.993	1.991	1.989	-
společnosti	14	1.613	1.530	1.715	1.594	1.579	1.451	1.667	1.817	1.814	1.822	1.896	1.818	1.666	1.533	1.758	1.807
národní	13	1.983	1.895	1.752	1.872	1.839	1.839	1.817	1.814	1.822	1.896	1.918	1.876	1.927	1.890	-	-
socialistického	12	1.303	1.293	1.298	1.297	1.292	1.295	1.295	1.297	1.293	-	1.292	1.292	1.291	-	-	-
světě	12	1.606	-	1.737	1.683	1.732	-	1.790	1.806	1.845	1.872	1.774	1.660	1.611	-	-	1.822
Československa	12	1.975	1.988	1.979	1.974	1.996	1.982	1.975	1.975	-	1.983	1.978	1.979	1.976	-	-	-
dráží	11	1.097	1.095	1.098	1.098	-	1.097	1.098	1.098	1.097	1.098	-	1.098	1.098	-	-	-
miru	11	1.896	-	-	1.891	1.908	1.907	1.894	1.899	1.942	1.944	1.905	1.933	1.897	-	-	-
příče	11	1.715	1.325	1.619	1.662	-	1.520	1.741	1.746	-	1.494	1.371	1.755	-	-	-	1.584
příčci	11	1.495	1.732	1.798	1.865	1.653	1.544	1.603	-	1.759	-	1.640	-	-	-	1.857	1.812
strany	11	1.733	1.830	1.852	1.554	1.801	1.616	1.729	1.823	-	-	1.698	1.717	1.579	-	-	-
vlasti	11	-	1.973	1.977	1.977	1.991	1.976	1.972	1.979	1.984	1.981	1.986	1.969	-	-	-	-
nového	10	-	1.766	-	1.832	1.728	1.724	-	1.815	1.823	1.830	1.826	1.735	1.768	-	-	-
roce	10	1.681	-	1.437	1.554	1.573	1.419	1.611	1.560	1.505	-	1.493	-	-	1.359	-	-
rok	10	1.715	1.709	1.699	1.827	-	-	1.447	1.664	1.678	-	1.610	1.527	1.659	-	-	-
jménem	9	1.881	-	-	1.854	-	1.876	1.859	1.891	1.922	1.900	-	1.928	1.863	-	-	-
kommunistické	9	1.913	1.980	1.979	-	1.979	-	1.988	1.984	-	-	1.972	1.966	1.964	-	-	-
výsledky	9	1.906	-	-	1.883	1.812	-	1.889	1.833	1.840	1.884	1.805	1.817	-	-	-	-
život	9	-	1.432	1.612	1.677	1.490	1.483	1.631	1.542	-	-	1.590	-	1.435	-	-	-
budeme	8	1.799	-	1.826	1.661	1.713	-	-	1.744	-	-	1.765	1.921	1.874	-	-	-
fronty	8	-	1.988	1.986	1.987	-	1.972	1.988	-	-	-	1.979	1.979	1.969	-	-	-
další	7	1.172	-	-	-	1.110	1.245	-	1.191	-	-	1.230	1.130	-	-	-	1.207
dalších	7	-	1.524	-	-	-	1.730	-	-	-	1.861	1.559	1.586	1.527	1.699	-	-
hospodářství	7	1.362	-	1.948	-	1.908	-	1.972	-	1.958	-	1.945	-	1.970	-	-	-
jame	7	-	1.310	1.174	1.108	1.336	1.219	1.439	-	-	-	-	0.885	-	-	-	-
přátelē	7	-	-	1.935	1.921	1.934	-	1.924	-	1.958	1.968	-	-	-	1.955	-	-
rokem	7	1.885	1.885	1.901	-	1.899	1.897	1.883	-	-	1.938	-	-	-	-	-	-
rozvoje	7	-	1.932	-	1.946	-	1.956	-	1.998	-	-	1.937	1.941	-	-	-	1.961
dále	6	-	1.716	1.752	-	1.795	1.743	-	1.817	-	1.728	-	-	-	-	-	-
pracujících	6	1.368	-	-	1.367	1.362	-	-	-	-	1.361	1.368	-	1.368	-	-	-
státy	6	1.898	-	-	-	1.954	1.897	-	-	1.967	-	-	1.935	-	-	-	1.942
životā	6	1.526	-	1.782	1.771	-	-	1.805	1.623	1.723	-	-	-	-	-	-	-
dobrē	5	1.857	1.804	1.724	-	-	-	1.677	-	1.760	-	-	-	-	-	-	-
mezinārodní	5	-	-	-	-	1.852	-	-	-	-	-	1.909	1.906	1.913	1.875	-	-
mūžeme	5	1.778	1.646	-	1.713	-	-	-	-	-	-	1.749	-	-	1.730	-	-
nārodū	5	1.973	1.961	1.966	-	-	-	1.970	1.969	-	-	-	-	-	-	-	-
rozvoj	5	1.878	-	1.947	-	-	1.891	-	-	-	1.913	1.932	-	-	-	-	-
soudruzi	5	-	1.988	-	1.990	-	-	-	1.990	-	-	1.993	1.989	-	-	-	-
soudruzi	5	-	1.998	-	1.999	-	-	-	1.999	-	-	1.999	1.999	-	-	-	-
soudruzi	5	1.999	-	-	-	-	-	1.999	-	-	-	1.999	1.999	-	-	-	1.999



MD-CADS: práce s časovou osou



Diachronní změna v KWA

Hypotézy

- ▶ různá sada KWs v různých letech indikuje změny ve společnosti/politice (viz výše)
- ▶ v čase se mění zapojení klíčového slova do kontextu ostatních klíčových slov naznačuje změnu asociací
- ▶ „collocations create connotations“: různé kontexty klíčových slov v čase naznačují proměny významu



Klíčové slovo *komunistické* / KSC u GH

Rok	Klíčové slovo	Fq	Levý kontext*
1975	<i>komunistické</i>	5	jménem, orgány strany, správná politika
1976	<i>komunistické</i>	6	jménem, orgány strany, úspěšná politika
1977	<i>komunistické</i>	5	jménem, orgány strany, podpora politice
1978	–	–	–
1979	<i>komunistické</i>	5	jménem, orgány strany, podpora politice
1980	–	–	–
1981	<i>komunistické</i>	4	jménem, orgány strany, kritika
1982	<i>komunistické</i>	5	jménem, orgány strany, podpora politice
1983	KSC	5	jménem, orgány strany, správná politika
1984	–	–	–
1985	<i>komunistické</i>	4	jménem, orgány strany, důvěra v politiku
1986	<i>komunistické</i>	3	jménem, orgány strany
1987	<i>komunistické</i>	3	jménem, orgány strany, KSSS
1988	KSC	3	orgány strany
1989	KSC	4	orgány strany

* V pravém kontextu je vždy slovo *strany*.





Vliv referenčního korpusu



Vliv referenčního korpusu

Jakým způsobem funguje referenční korpus v KWA?

velikost: čím větší korpus (nebo text), tím víc KWs

složení: různé referenční korpusy reprezentují různé modelové čtenáře

- ▶ vyvážený korpus ~ průměrný čtenář
- ▶ specializovaný korpus ~ specifický čtenář (např. se specifickým odborným zázemím...)



Různí příjemci = různé interpretace

Kontrastivní KWA

- ▶ různé referenční korpusy: interference – čas, styl, téma
- ▶ Novoroční projevy Gustáva Husáka (1975–1989)
- ▶ současný čtenář (SYN2010) × „čtenář z minulosti“ (Totalita)



Analýza GH: vliv referenčního korpusu

Jak se změní výsledky KWA s různými referenčními korpusy?

- ▶ inventář KWs zůstává více méně stejný
- ▶ liší se jejich pořadí (prominence – DIN)

Historický čtenář (Totalita)

- žánrové rozdíly
- ▶ modální slovesa: chtít, moci
- ▶ Slovesa: 1. sg./pl.

Současný čtenář (SYN2010)

- KWs spojená s historickými reáliemi
- ▶ ideologická slova
- ▶ archaismy, historismy



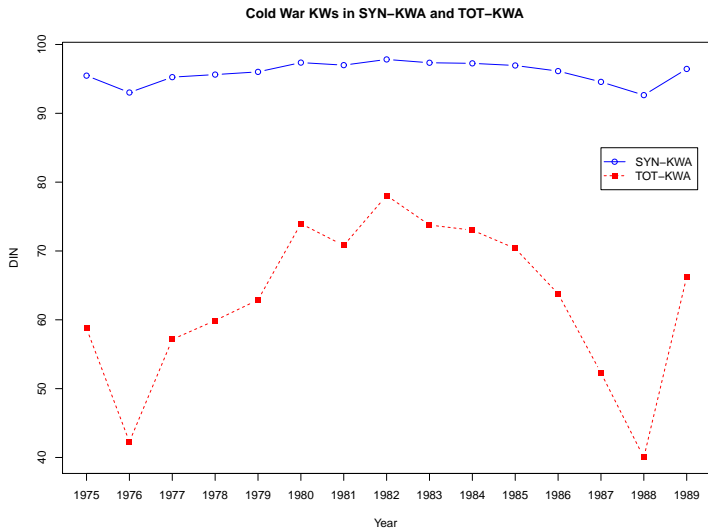
Detailnější srovnání – 3 skupiny slov

Cold war: mír, míru, mírová, mírové, mírového, mírovému, mírovou, mírový, mírových, mírovými, mírumilovné, mírumilovných, mírumilovným; napětí; odzbrojení, výzbroje, zbrojení, zbrojením, ozbrojených

Collective possession: náš, naše, našeho, našem, našemu, naši, naší, našich, našim, naším, našimi

Ideo markers: socialismu, socialismus, socialistická, socialistické, socialistického, socialistickém, socialistickému, socialistickou, socialistický, socialistických, socialistickým, socialistickými; komunismu, komunisté, komunistů, ksč, komunistům, komunisty, komunistická, komunistické, komunistickým

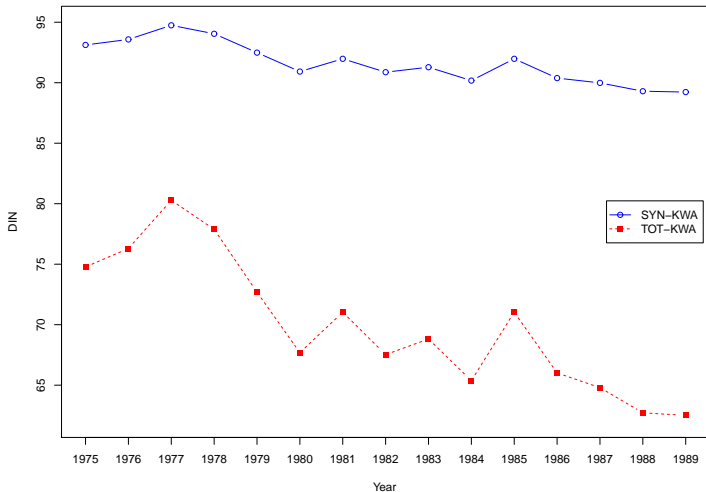
Cold war



Fidler-Cvrček (2015)

Collective possession

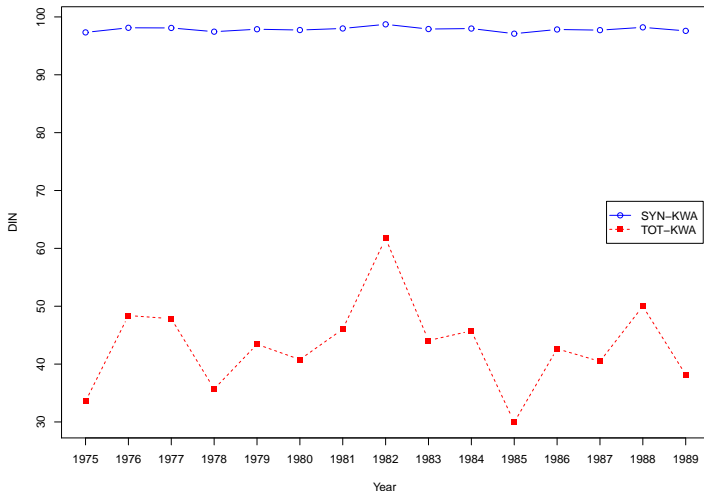
KWs "our" in SYN-KWA and TOT-KWA



Fidler-Cvrček (2015)

Ideological markers

Ideological markers KWs in SYN-KWA and TOT-KWA



Fidler-Cvrček (2015)

Rozdíl je v citlivosti

- ▶ vývoj prominence jednotlivých témat je principiálně stejný
- ▶ současný čtenář má celkově vyšší míru prominence (DIN)
- ▶ vývojové tendence jsou zřetelnější pro čtenáře z minulosti
- ▶ současný čtenář nedokáže vnímat rozdíly





Morfologie a prominence



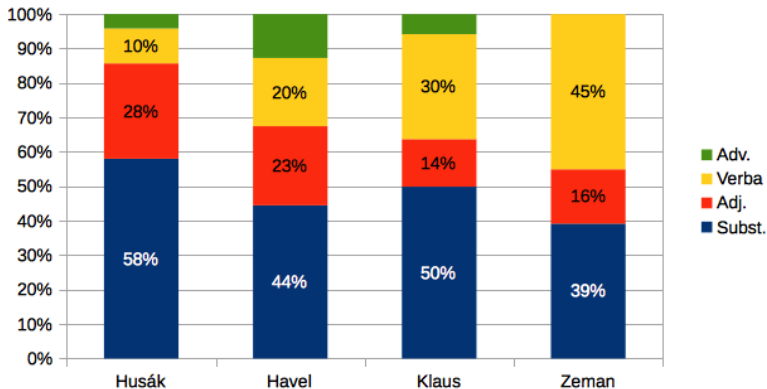
Řečnické strategie

Zaměření na gramatické a flektivní charakteristiky KWs

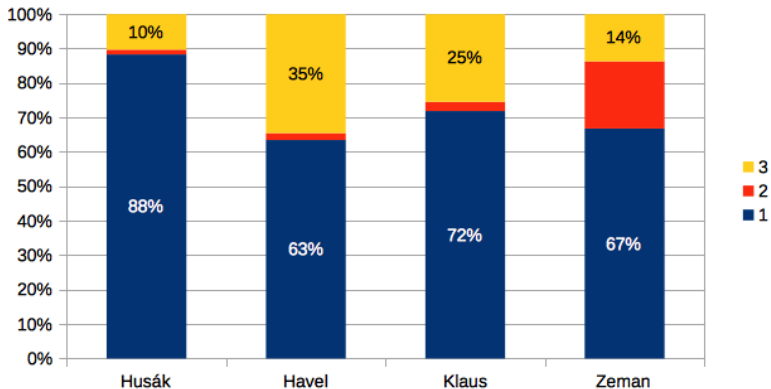
- ▶ porovnání prominence gramatických kategorií v projevech 4 prezidentů: G. Husák, V. Havel, V. Klaus a M. Zeman
- ▶ pomocí DIN se zohledňuje působení na současného nativního mluvčího (referenční korpus SYN2010)
- ▶ zastoupení gramatických kategorií v rámci klíčových slov – obvyklé gramatické charakteristiky prominentních slov



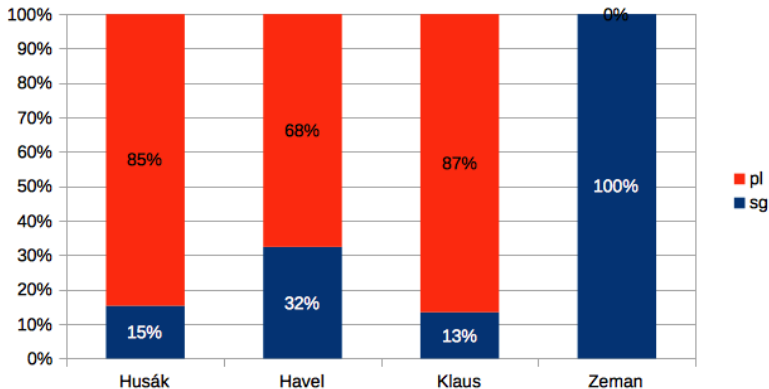
Slovní druhy



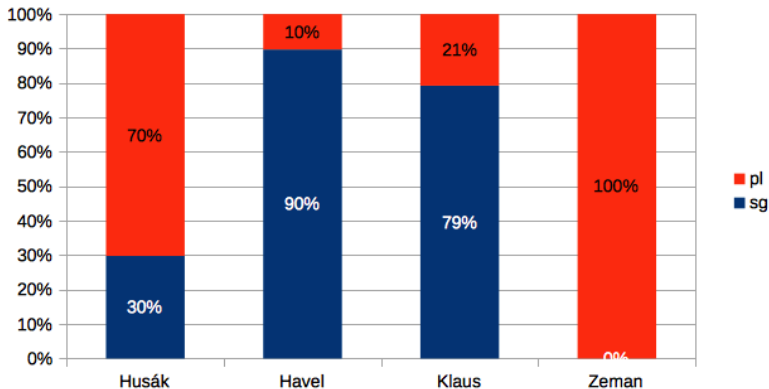
Slovesná osoba




Distribuce čísla – 1. os.



Distribuce čísla – 3. os.



 Tvary nebo lemmata?



Analýza portálu Sputniknews.cz

Data

- ▶ Korpus SPU
 - ▶ texty obsahující slovní základ *rusk-*
 - ▶ z období 4.3. – 4.4. 2015
 - ▶ 80 000 pozic
- ▶ Referenční korpus: SYN2015
 - ▶ reprezentativní korpus současné psané češtiny (FIC, NFC, NWS)
 - ▶ z období 2010–2014
 - ▶ 121 666 414 pozic

Iniciální analýza byla provedena na lemmatech (MorphoDiTa).



Prominentní lemmata

- ▶ přes 500 klíčových slov, na špičce prominence propria
- ▶ lemmata inherentně poukazující na specif. zájmy Sputniku:
 - ▶ *rusofobie* (37.)
 - ▶ *protiruský* (53.)
- ▶ lemmata, jejichž užití vysvětluje z kontextu
 - ▶ *kompromitace* (70.) – Lavrov označil verzi Reuters o výbuchu malajského Boeingu za lživou **kompromitaci**.
 - ▶ *rozmístování* (78.) – Bude-li **rozmístování** PRO u ruských hranic pokračovat, bude zcela jasné, že ve skutečnosti se předpokládá využít těchto systémů proti Rusku.
 - ▶ *démonizace* (81.) – Peskov hovořil o příčinách **démonizace** Putina na Západě.



Rusko a jeho pády

Lemma *Rusko* je prominentní – dáno metodou sběru dat; liší se ovšem prominencí jednotlivých pádů

Pád	Tvar	SPU	SYN2015	DIN
Nom.	<i>Rusko</i>	175	2088	7.94
Gen.	<i>Ruska</i>	164	2351	-1.22
Dat.	<i>Rusku</i>	93	493	45.04
Ak.	<i>Rusko</i>	60	546	21.18
Lok.	<i>Rusku</i>	53	2457	-53.64
Inst.	<i>Ruskem</i>	66	613	20.20
SUM		611	8548	

DIN počítán vzhledem k celkovým frekvencím lemmatu *Rusko* v obou korpusech.

Předložky s dativem *Rusku*

Prep. + <i>Rusku</i>	SYN2015		SPU	
	AbsFq	RelFq	AbsFq	RelFq
k	130	40.1 %	18	23.7 %
proti	105	32.4 %	43	56.6 %
kvůli	8	2.5 %	0	0.0 %
díky	0	0.0 %	0	0.0 %
vůči	79	24.4 %	15	19.7 %
oproti	1	0.3 %	0	0.0 %
navzdory	0	0.0 %	0	0.0 %
naproti	0	0.0 %	0	0.0 %
vstříc	0	0.0 %	0	0.0 %
blízko	0	0.0 %	0	0.0 %
napospas	1	0.3 %	0	0.0 %



Interpretace

Jak interpretovat prominenci Dat., Ak. a Lok. lemmatu *Rusko*?

- ▶ Rusko ve SPU není prezentováno jako agens, ale jako patiens
- ▶ (zejm. s ohledem na prominenci dat. s předložkou *proti*): Rusko je zobrazováno jako cíl nebo oběť
- ▶ může být důsledkem záměru vzbudit v čtenáři sympatie k Rusku





Závěr



...so what?

K čemu je empirická analýza diskurzu, odhaluje-li věci intuitivně tušené?

- ▶ empirické ověřování hypotéz je základem vědeckého postupu
- ▶ empirie pomáhá odhalovat možné příčiny „tušení“
- ▶ bez empirického zkoumání textů nejsme schopni simulovat interpretace jiných lidí
- ▶ ...
- ▶ ne vše, co je intuitivně tušené, se empirickým testováním potvrdí
- ▶ ne vše je tušené (existují i *wow!* závěry)



Těším se na vaše otázky

