ČESKÝ NÁRODNÍ
KORPUS

ČESKÝ NÁRODNÍ KORPUS

Dějiny v zrcadle jazyka

Václav Cvrček, ÚČNK

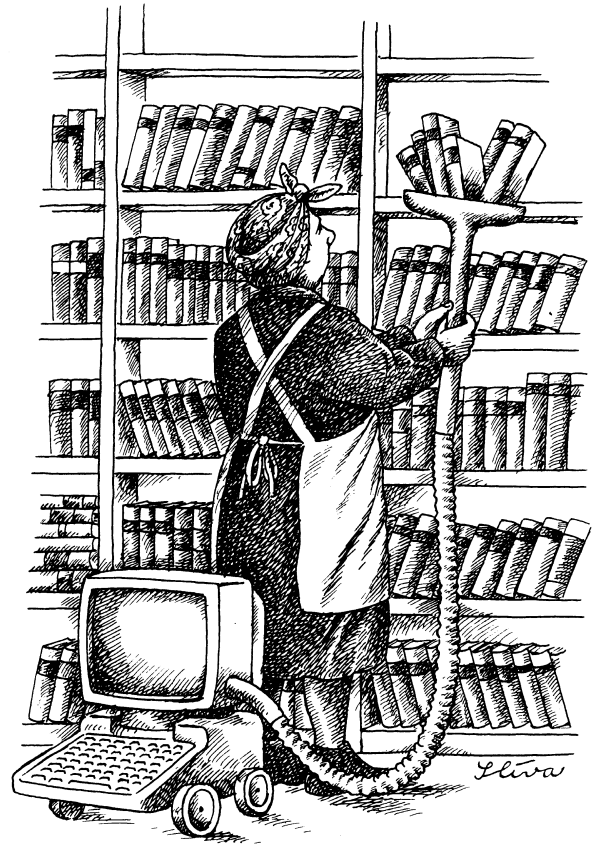
24. února 2015

Jazykové korpusy

Co je jazykový korpus?

Co je korpus?

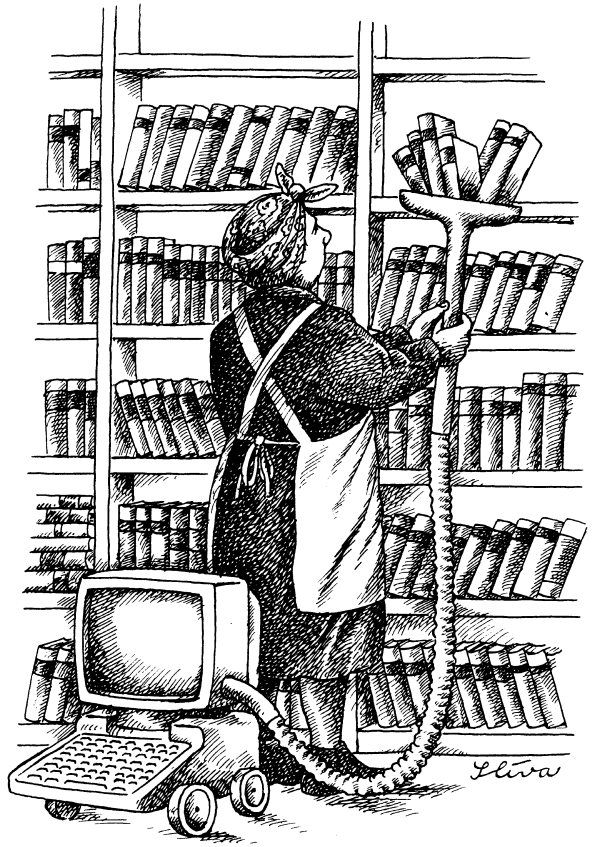
- databáze autentických textů/promluv



Co je jazykový korpus?

Co je korpus?

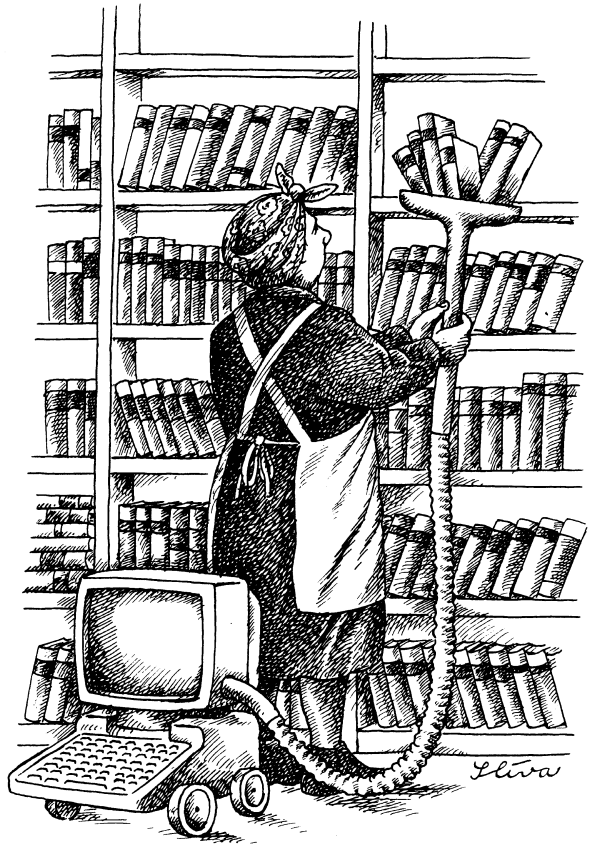
- databáze autentických textů/promluv
- uložená elektronicky v jednotném formátu



Co je jazykový korpus?

Co je korpus?

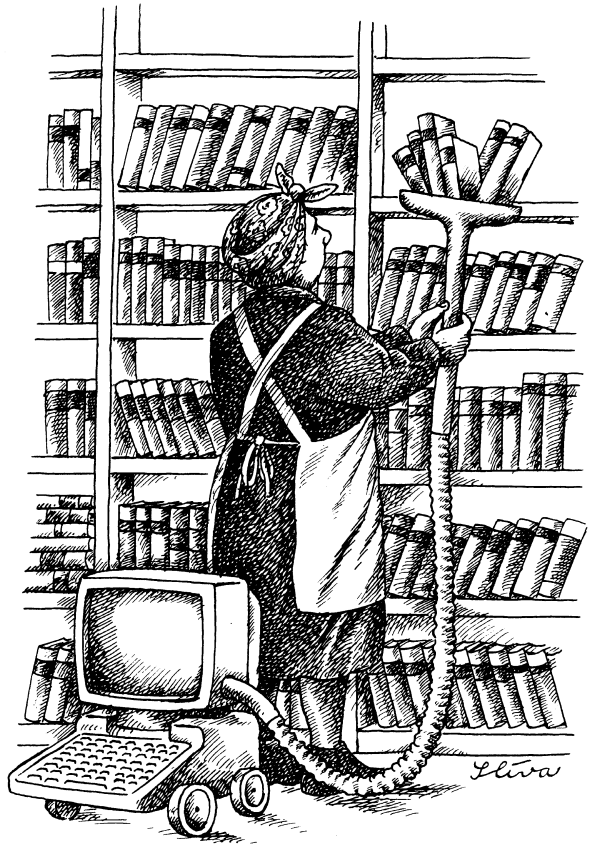
- databáze autentických textů/promluv
- uložená elektronicky v jednotném formátu
- prohledávatlená



Co je jazykový korpus?

Co je korpus?

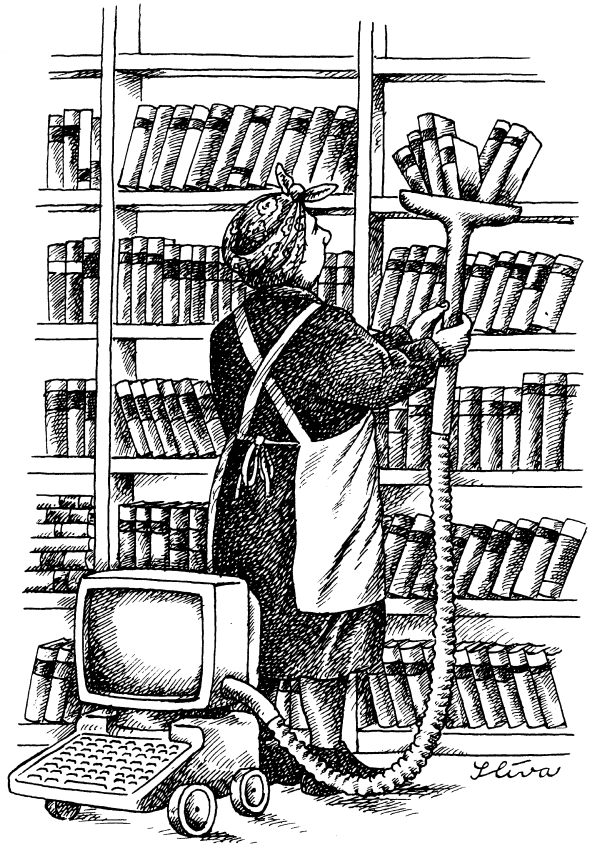
- databáze autentických textů/promluv
- uložená elektronicky v jednotném formátu
- prohledávatlená
- reprezentativní vzhledem ke svému účelu



Co je jazykový korpus?

Co je korpus?

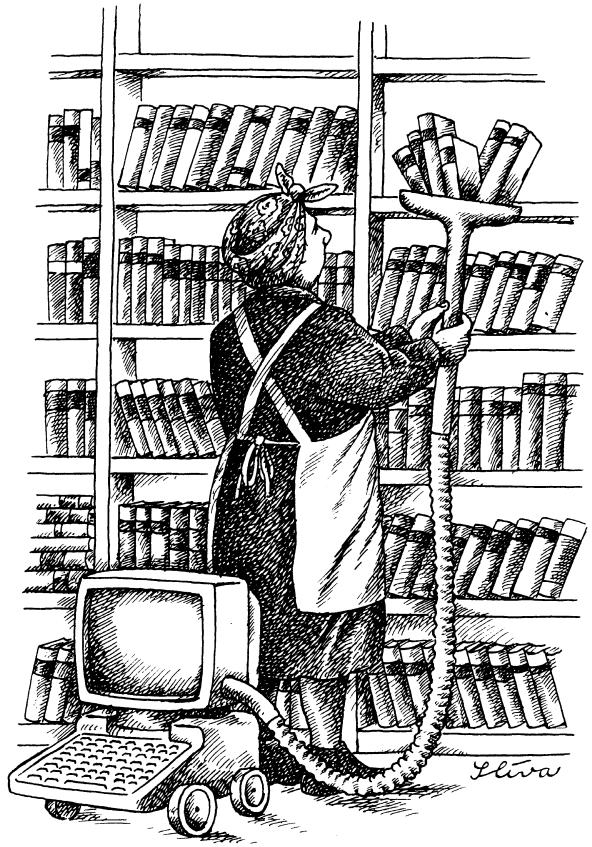
- databáze autentických textů/promluv
- uložená elektronicky v jednotném formátu
- prohledávatlená
- reprezentativní vzhledem ke svému účelu
- zobrazující jevy v jejich přirozeném a typickém kontextu



Co je jazykový korpus?

Co je korpus?

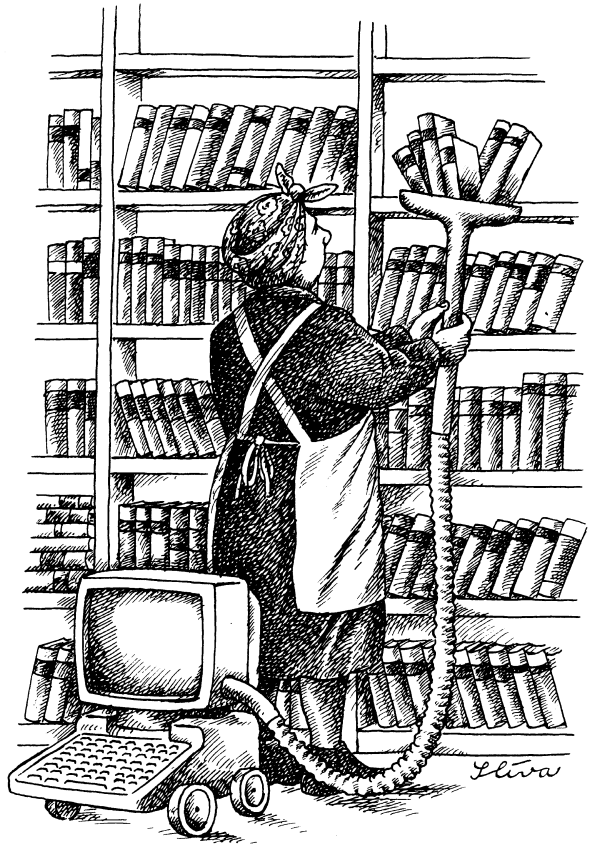
- databáze autentických textů/promluv
- uložená elektronicky v jednotném formátu
- prohledávatlená
- reprezentativní vzhledem ke svému účelu
- zobrazující jevy v jejich přirozeném a typickém kontextu
- spolu s informací o frekvenci



Co je jazykový korpus?

Co je korpus?

- databáze autentických textů/promluv
- uložená elektronicky v jednotném formátu
- prohledávatlená
- reprezentativní vzhledem ke svému účelu
- zobrazující jevy v jejich přirozeném a typickém kontextu
- spolu s informací o frekvenci
- původně pro lexikografické účely

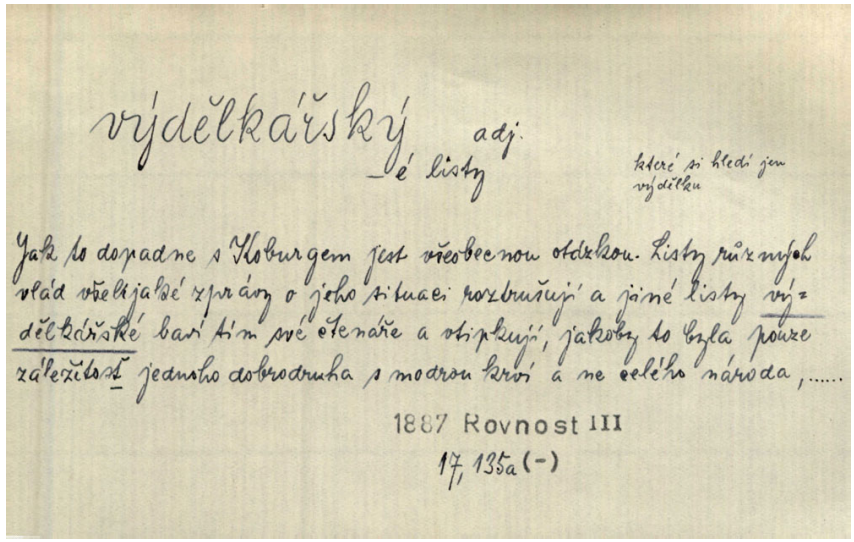


Konkordance a KWIC

Mladá fronta DNES, 10. 9. 2008	lesy , které stále zabírají třetinu země " . Lonely Planet také hodně stojí na	kuriozitách	. Takže jeho autorům neušel Tábor , kam zve speciální rámeček věnovaný " náboženské sektě
Právo, 2. 6. 2006	. Jihočeští hasiči lovili otužilce i divočáka Kromě běžných zásahů zažívají jihočeští hasiči i opravdové	kuriozity	. Lidé jim volají , aby vytáhli z vody topícího se člověka , a když
Sedmička, č. 18/2009	německy . Text byl adresován Tutzi , tedy šlechtičně Terezii Seilern-Aspangové , roz. Lažanské .	Kuriozitou	byl také zajímavý doklad , rovněž ze září 1943 , představující soupis zvěře určené k
Lidové noviny, 3. 1. 2006	instalaci současného umění . Díla mezinárodní provenience Jakkoli se artefakty tváří jako průmyslové zboží ,	kuriozity	nebo upomínkové předměty , svůj charakter výtvarného díla nepopírají . Jména autorů uvádějí coby výrobce
ŠíP, 17. 4. 2007	v Česku namalovaný pastelkami . Včera ho v Pelhřimově zaregistrovali do české databanky rekordů a	kuriozit	. Autorkou je výtvarnice (17) , studentka umělecké školy v Jihlavě Eva Toulová
Nezbytné věci	plechovka byla pochopitelně Todda - toho Todda , který by nepochybně chtěl tábořit v obchodě	Kuriozity	Auburn , kdyby mu to dovolili . Byl uchvacován těmi pouťovými blbostmi , které tam
Metro, 21. 2. 2008	která bude vozit české fanoušky na červnové fotbalové mistrovství Evropy v Rakousku a Švýcarsku .	KURIOZITA	. Lokomotivu řady Taurus , kterou vyrábí firma Siemens a jež drží světový rekord v
Masožravé rostliny	rosnatka stínomilná (foto 217 - 219) Výjimečně široké listy řadí tuto rosnatku mezi	kuriozity	rodu . Každý pěstitel by ji chtěl mít ve sbírce , avšak leckdy se nedaří
Právo, 14. 8. 2008	Letošní doplňovací volby do horní parlamentní komory za volební obvod 51 - Žďársko provází skutečná	kuriozita	: proti sobě v nich stojí dva kandidáti , kteří navzájem dlouhodobě zaměstnávají orgány činné
Hospodářské noviny, 2. 3. 2006	i znepokojením sledují , že ropuchy v Austrálii začínají mít delší nohy . Není to	kuriozita	, ale ukázka rychlého evolučního vývoje . Pro australské životní prostředí je nepříjemné , že
Pevné pouto	zlehka na okázalém škrabáku na boty , který babička Lynn nalezla v jednom obchodě s	kuriozitami	v Marylandu . Buckley házel pískající gumovou kravičkou , zatímco Holiday pro ni běhal a
Deníky Moravia, 19. 11. 2005	knihy , fotoaparáty , obrazy , hodiny , hračky , sklo , porcelán a další	kuriozity	. Akce se koná od 7 do 12 hodin . Vozy značky Lancia budou mít
Mladá fronta DNES, 31. 12. 2005	představitelé krajského úřadu . Práci s jejich vyhledáváním si dali pracovníci odboru cestovního ruchu .	Kuriozity	jsou totiž podle nich jedním z největších lákadel pro turisty . Seznam všech nej v
Sny robotů	- , a on tvrdil , že pomocí téhle metody se antigravitace stane nejen vědeckou	kuriozitou	, ale praktickou , průmyslově využitelnou záležitostí . Snad . Ale bylo to nedokonalé o
Mladá fronta DNES, 28. 1. 2008	se stal právě pohyb . Na čtyřicet výtvarníků je dostalo do České knihy rekordů a	kuriozit	jako nejdelší obraz v Česku . Na rekord dohlížela Agentura Dobrý den z Pelhřimova .
Nedělní svět, č. 2/2006	psal v letech 1965 až 1992 pro česká nakladatelství . Jeho trefné postřehy nejsou pouhou	kuriozitou	. Ukazují ojedinělou schopnost vměstnat i na drobný prostor neotřelé argumenty , s nimiž se
Mladá fronta DNES, 24. 10. 2008	dny ponesou v ryze americkém duchu , do Větřkovic na Opavsku vyrazí sběratelé hlavolamů a	kuriozit	z různých koutů Česka i ze zahraničí . Ve Vlčovicích na Novojičínsku si zase lidé
Lidé a země, č. 11/2005	. Ilustrace , sošky , zápisy v kronikách i řada dalších , často velice podivuhodných	kuriozit	, vypráví o velké lásce , kterou Švýcaři ke svému největšímu hrdinovi chovají . Z
Hospodářské noviny, 10. 4. 2006	Objeví i rozsáhlý soubor pivních korbelů z materiálů všeho druhu a v neposlední řadě expozici	kuriozit	. Sem patří nejmenší soudek na světě , džbánek měřící jeden centimetr nebo obří korbel
Lidé a země, č. 3/2006	a zpětně proudící řeka se sama postará o bezpracný návrat . Tohle je však spíše	kuriozita	. Všechny ostatní kánoe , co potkáváme , jsou poháněny pádly . Asmaté tradičně pádlují
Deníky Bohemia, 12. 4. 2007	kola z automobilů nebo vrtačky a brusky , " řekla Hábrová . Mízí ale i	kuriozity	. V únoru z jednoho lochotínského sklepa odnesl zloděj kompletní zbroj historického šermu . "
Týdeník Květy, č. 13/2009	" . Nesympatickou léčebnou pomůcku drží na zádech tři andělíčci a šťastně se usmívají .	Kuriozity	se nevyhýbají Moskvě - lze tam spatřit památník prvnímu taburetu Ruska , tavenému sýru Družba

Předchůdci korpusu – lístkový katalog

Oddělení současné lexikologie a lexikografie, ÚJČ AV ČR (od r. 1891)



http://lexiko.ujc.cas.cz

Lístkový katalog

Lístkový katalog

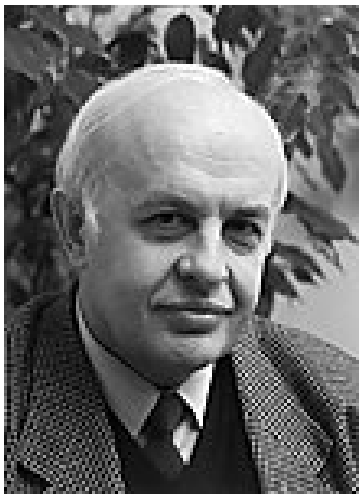
- výběrový (totální excerpce jen na malé části – Mácha, Erben, Němcová, Neruda, Vrchlický, Jirásek)
- problematická reprezentativnost
- stáří materiálu
- pořizovací náklady
- nepřehlednost
- omezený kontext

Brown corpus (1964)

Henry Kučera
(1925–2010)

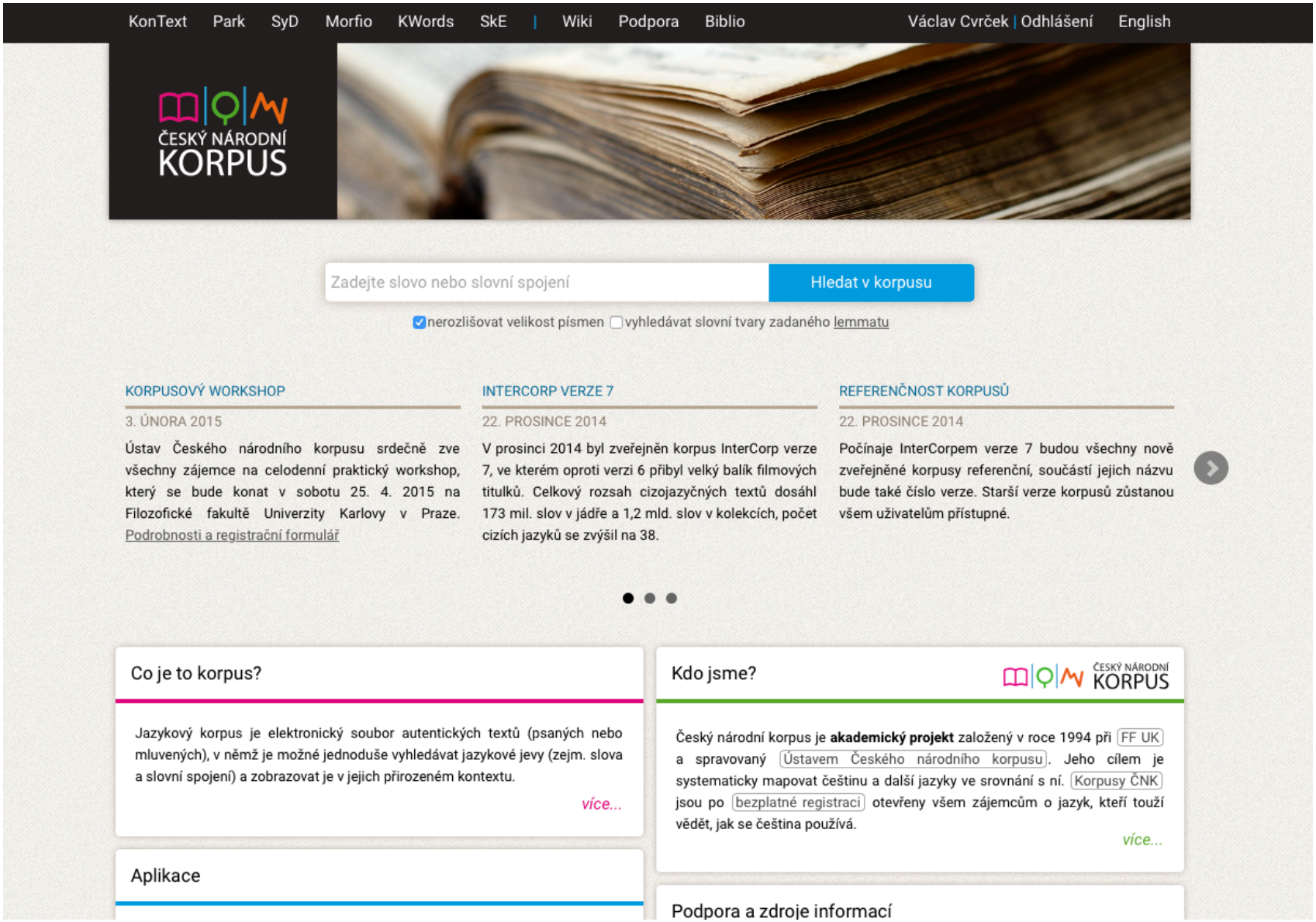
- H. Kučera – N. Francis: *Brown University Standard Corpus of Present-Day American English*
- první elektronický korpus
- 500 vzorků po 2000 slovech z publikací z roku 1961
- sofistikovaný design, různé žánry a textové typy
- *American Heritage Dictionary* – počítačově zpracovaný frekvenční slovník
- spell checker

Český národní korpus

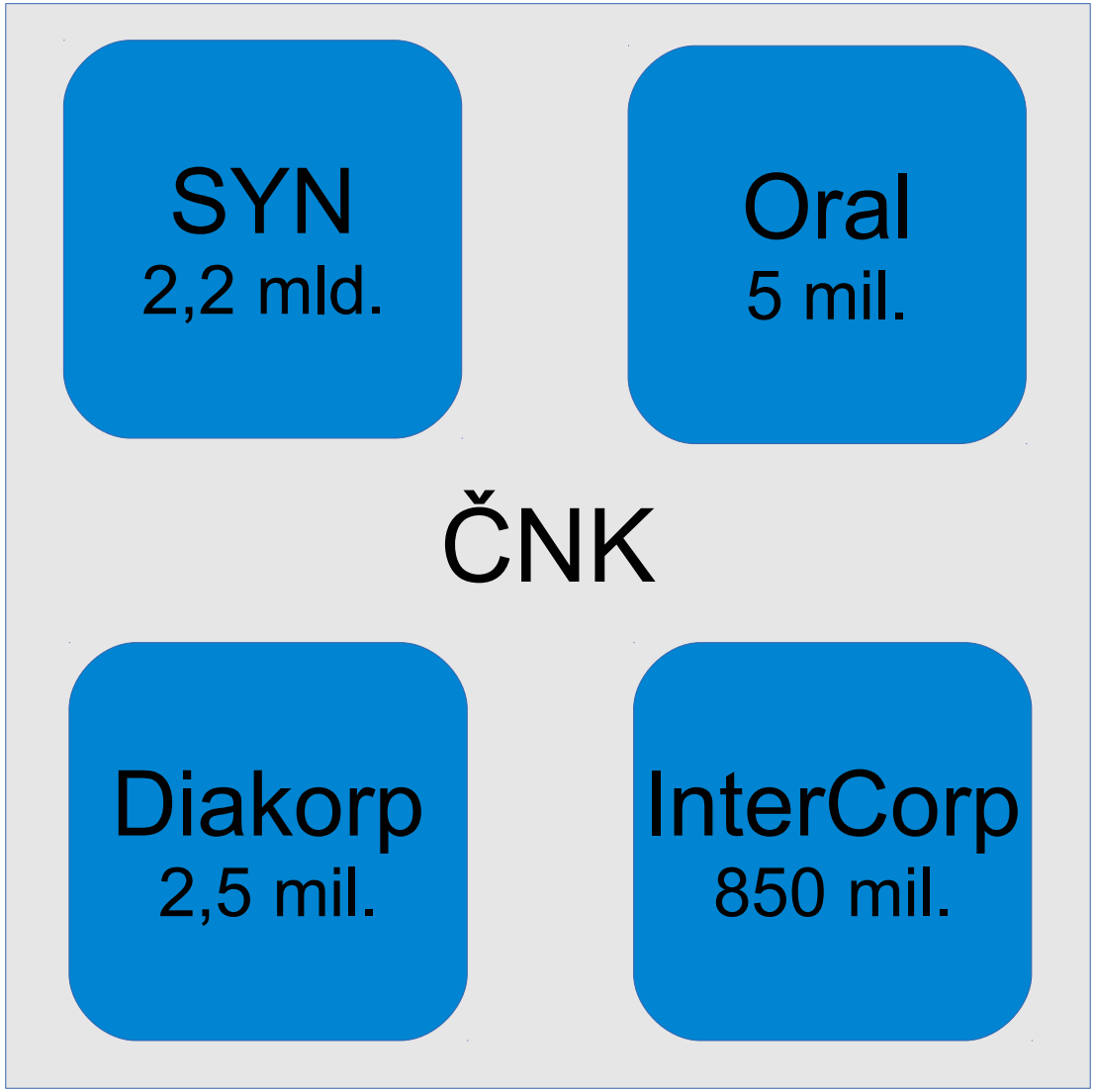


František Čermák

- založen v roce 1994
- Ústav Českého národního korpusu FF UK
- veřejně přístupné, zdarma (po registraci)
- spolupráce s ÚTKL, ÚFAL MFF ad.
- obor matematická lingvistika
- Portál ČNK – http://www.korpus.cz


KonText Park SyD Morfio KWords SkE | Wiki Podpora Biblio
Václav Cvrček | Odhlášení English
ČESKÝ NÁRODNÍ KORPUS
Zadejte slovo nebo slovní spojení
Hledat v korpusu
nerozlišovat velikost písmen
vyhledávat slovní tvary zadaného lemmatu
KORPUSOVÝ WORKSHOP
3. ÚNORA 2015
Ústav Českého národního korpusu srdečně zve všechny zájemce na celodenní praktický workshop, který se bude konat v sobotu 25. 4. 2015 na Filozofické fakultě Univerzity Karlovy v Praze. Podrobnosti a registrační formulář
INTERCORP VERZE 7
22. PROSINCE 2014
V prosinci 2014 byl zveřejněn korpus InterCorp verze 7, ve kterém oproti verzi 6 přibyl velký balík filmových titulků. Celkový rozsah cizojazyčných textů dosáhl 173 mil. slov v jádře a 1,2 mld. slov v kolekcích, počet cizích jazyků se zvýšil na 38.
REFERENČNOST KORPUSŮ
22. PROSINCE 2014
Počínaje InterCorpem verze 7 budou všechny nově zveřejněné korpusy referenční, součástí jejich názvu bude také číslo verze. Starší verze korpusů zůstanou všem uživatelům přístupné.
Co je to korpus?
Jazykový korpus je elektronický soubor autentických textů (psaných nebo mluvených), v němž je možné jednoduše vyhledávat jazykové jevy (zejm. slova a slovní spojení) a zobrazovat je v jejich přirozeném kontextu.
více...
Kdo jsme?
ČESKÝ NÁRODNÍ KORPUS
Český národní korpus je akademický projekt založený v roce 1994 při FF UK a spravovaný Ústavem Českého národního korpusu. Jeho cílem je systematicky mapovat češtinu a další jazyky ve srovnání s ní. Korpusy ČNK jsou po bezplatné registraci otevřeny všem zájemcům o jazyk, kteří touží vědět, jak se čeština používá.
více...
Aplikace
Podpora a zdroje informací

Struktura ČNK



ČNK: Řada SYN

Korpusy současného psaného jazyka

- korpusy reprezentativní: SYN2000, SYN2005, SYN2010 (100 mil. tokenů každý)

ČNK: Řada SYN

Korpusy současného psaného jazyka

- korpusy reprezentativní: SYN2000, SYN2005, SYN2010 (100 mil. tokenů každý)
- korpusy nereprezentativní: SYN2006PUB, SYN2009PUB a SYN2013PUB (300 mil., 700 mil. a 900 mil. tokenů)

ČNK: Řada SYN

Korpusy současného psaného jazyka

- korpusy reprezentativní: SYN2000, SYN2005, SYN2010 (100 mil. tokenů každý)
- korpusy nereprezentativní: SYN2006PUB, SYN2009PUB a SYN2013PUB (300 mil., 700 mil. a 900 mil. tokenů)
- dohromady: SYN (2,2 mld. tokenů)

ČNK: Řada SYN

Korpusy současného psaného jazyka

- korpusy reprezentativní: SYN2000, SYN2005, SYN2010 (100 mil. tokenů každý)
- korpusy nereprezentativní: SYN2006PUB, SYN2009PUB a SYN2013PUB (300 mil., 700 mil. a 900 mil. tokenů)
- dohromady: SYN (2,2 mld. tokenů)
- bibliografická informace, lingvistická anotace

ČNK: Řada SYN

Korpusy současného psaného jazyka

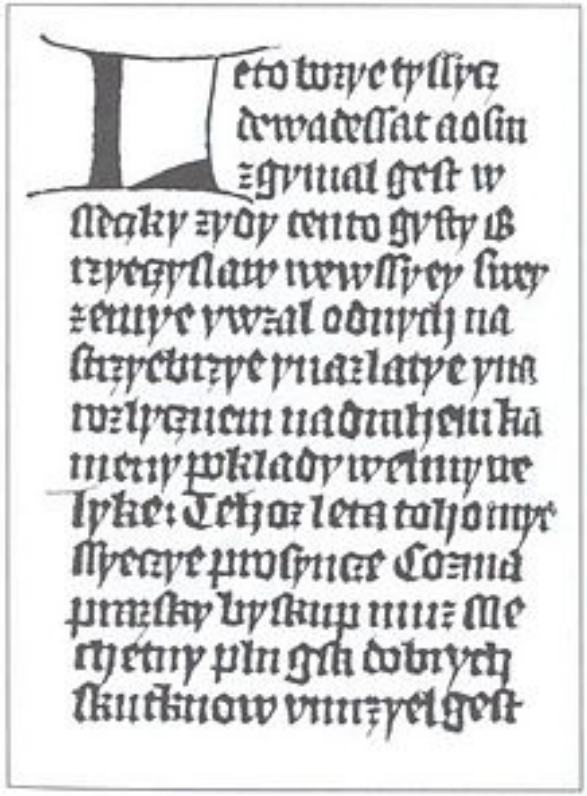
- korpusy reprezentativní: SYN2000, SYN2005, SYN2010 (100 mil. tokenů každý)
- korpusy nereprezentativní: SYN2006PUB, SYN2009PUB a SYN2013PUB (300 mil., 700 mil. a 900 mil. tokenů)
- dohromady: SYN (2,2 mld. tokenů)
- bibliografická informace, lingvistická anotace
- lemmatizace – přidělení základního slovníkového tvaru každému tokenu (*pána* $\rightarrow$ *pán*, *představuje* $\rightarrow$ *představovat*)

ČNK: Řada SYN

Korpusy současného psaného jazyka

- korpusy reprezentativní: SYN2000, SYN2005, SYN2010 (100 mil. tokenů každý)
- korpusy nereprezentativní: SYN2006PUB, SYN2009PUB a SYN2013PUB (300 mil., 700 mil. a 900 mil. tokenů)
- dohromady: SYN (2,2 mld. tokenů)
- bibliografická informace, lingvistická anotace
- lemmatizace – přidělení základního slovníkového tvaru každému tokenu (*pána* $\rightarrow$ *pán*, *představuje* $\rightarrow$ *představovat*)
- tagování – přidělení morfologické značky (tagu) každému tokenu (slovní druh, pád, číslo, rod, osoba atp.)

ČNK: Diakorp



Korpus starší psané češtiny

- 2.5 mil. slov (od 13. stol.)
- kde začíná jazyková současnost?
- problematika transkripce/transliterace – pravopisné reformy, jazykové změny
- automatické rozpoznávání textu (OCR) je nepoužitelné
- problém reprezentativnosti

Proč korpusy?

Empirický a teoretický výzkum jazyka

Teoretický přístup

+ využití introspekce/intuice
+ elegance popisu
+ snazší formalizovatelnost
- problém přiměřenosti

Empirický a teoretický výzkum jazyka

Teoretický přístup

+ využití introspekce/intuice
+ elegance popisu
+ snazší formalizovatelnost
- problém přiměřenosti

Empirický přístup

+ experiment založený na datech
+ statistické vyhodnocování
- pracné, nákladné
- nejednoznačné závěry

Korpus versus intuice

Doplňte typické *přídavné jméno* nebo *sloveso*

	Podle intuice	**Podle korpusu**
úsilí	_____	
snaha	_____	
výzva	_____	
problém	_____	

Korpus versus intuice

Doplňte typické *přídavné jméno* nebo *sloveso*

	Podle intuice	**Podle korpusu**
úsilí	_____	*vynaložený, vynaložit*
snaha	_____	*marný, udržet*
výzva	_____	*opakovaný, uposlechnout*
problém	_____	*zdravotní, řešit*

Korpus versus intuice

Doplňte typické *přídavné jméno* nebo *sloveso*

	Podle intuice	**Podle korpusu**
úsilí	_____	*vynaložený, vynaložit*
snaha	_____	*marný, udržet*
výzva	_____	*opakovaný, uposlechnout*
problém	_____	*zdravotní, řešit*

Seřaďte podle četnosti

- *obraz* –
- *stůl* –
- *židle* –

Korpus versus intuice

Doplňte typické *přídavné jméno* nebo *sloveso*

	Podle intuice	**Podle korpusu**
úsilí	_____	*vynaložený, vynaložit*
snaha	_____	*marný, udržet*
výzva	_____	*opakovaný, uposlechnout*
problém	_____	*zdravotní, řešit*

Seřaďte podle četnosti

- *obraz* – frekvence: 22 832; pořadí: 2.
- *stůl* – frekvence: 25 372; pořadí: 1.
- *židle* – frekvence: 7 668; pořadí: 3.

Jazyk jako zrcadlo společenských změn

Skupina A:
mobil, kraj, hejtman, kvůli, celebrita, developerský, webový, multifunkční, ustát, nalákat, zeleň; za volant, rodinný dům, obchodní centrum, letní posila, čerpání dotace…

Skupina B:
fax, ing, marka, úloha, výbor, tisk, tajemník, eventuální, západoevropský; sdělovací prostředky, cenný papír, liberalizace cen, pluralitní demokracie, národní majetek…

Jazyk jako zrcadlo společenských změn

Vývoj relativních frekvencí v letech 1992–2009

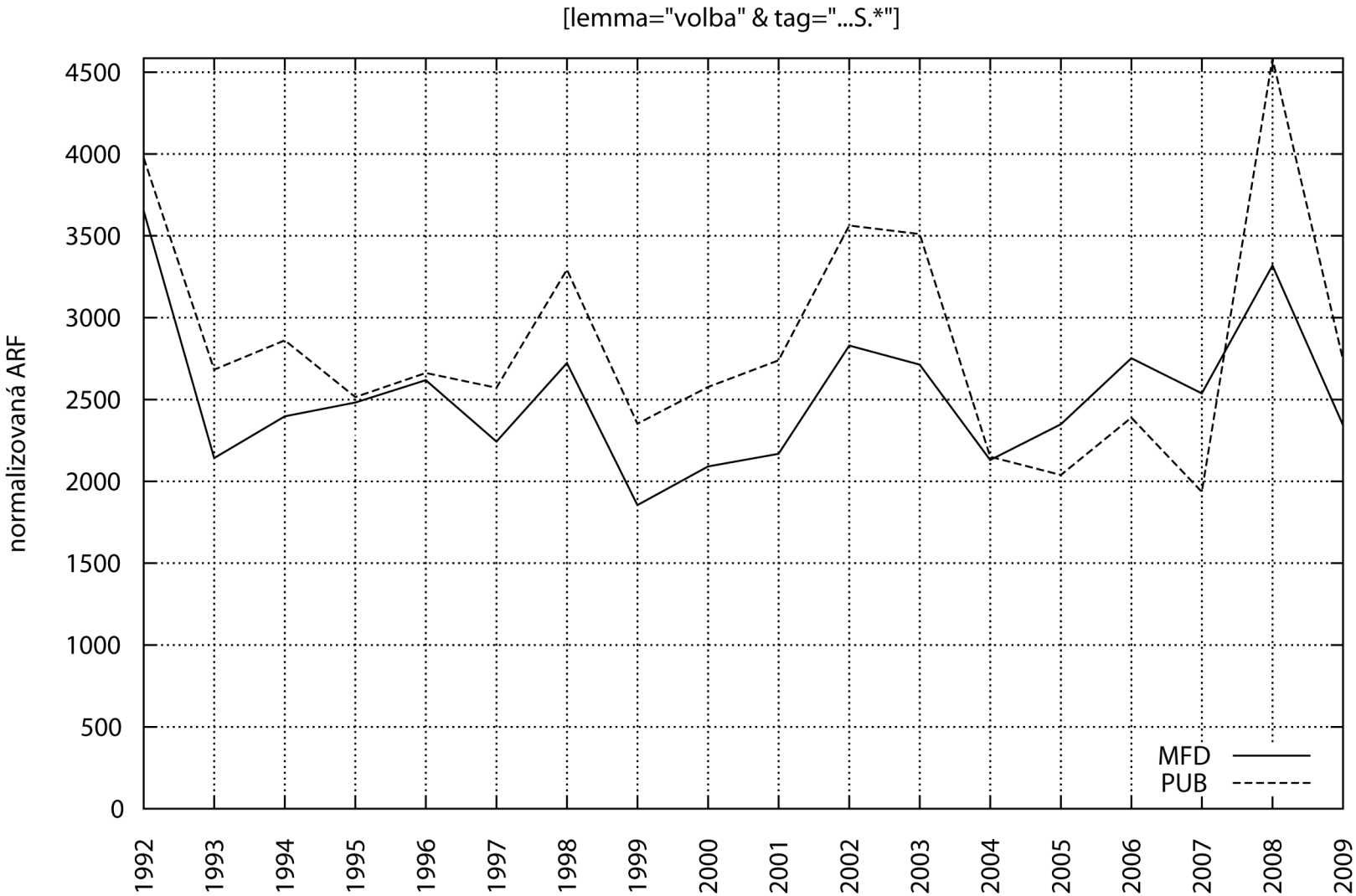
Skupina A:
mobil, kraj, hejtman, kvůli, celebrita, developerský, webový, multifunkční, ustát, nalákat, zeleň; za volant, rodinný dům, obchodní centrum, letní posila, čerpání dotace...

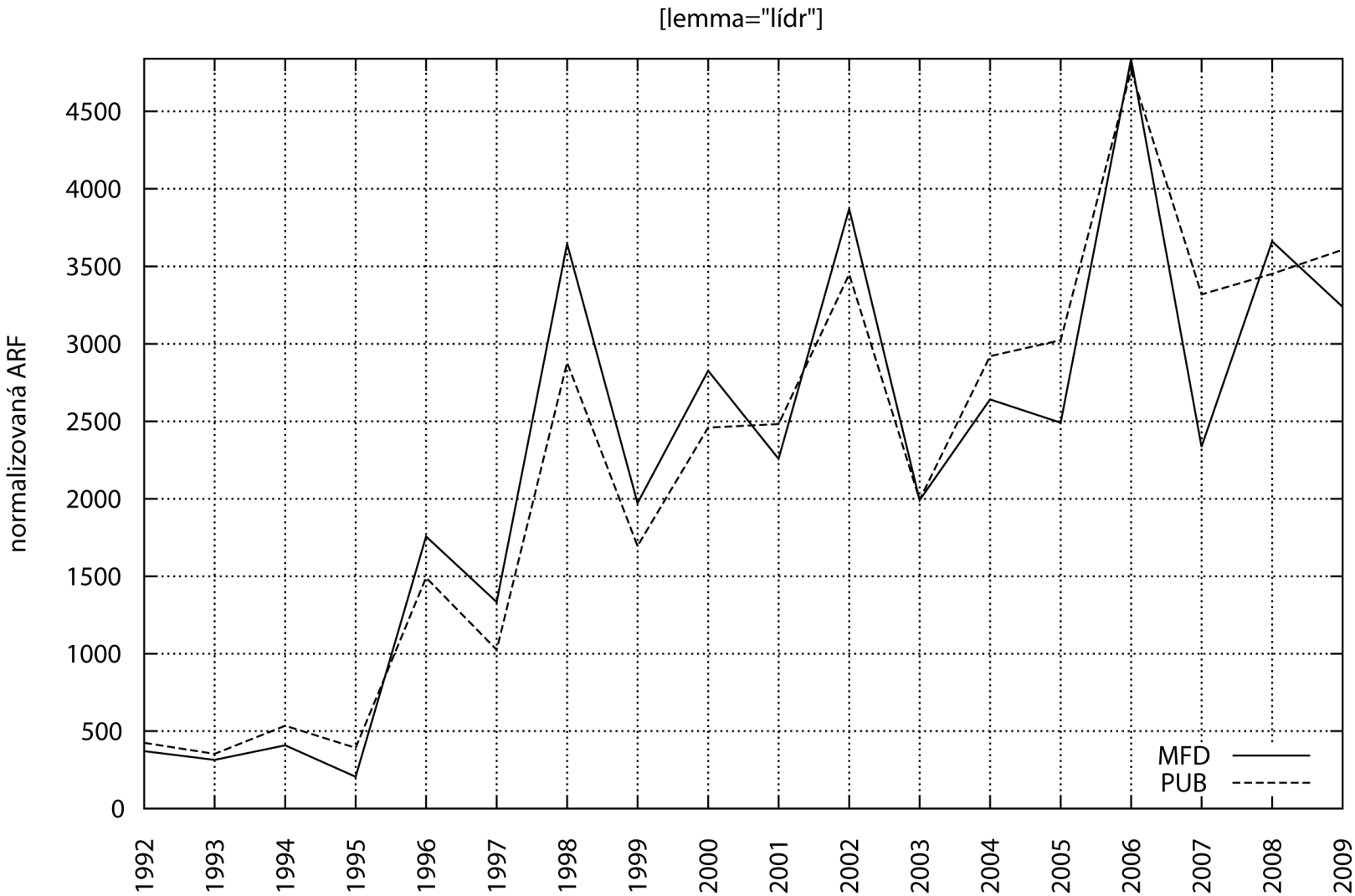
Trend rostoucí

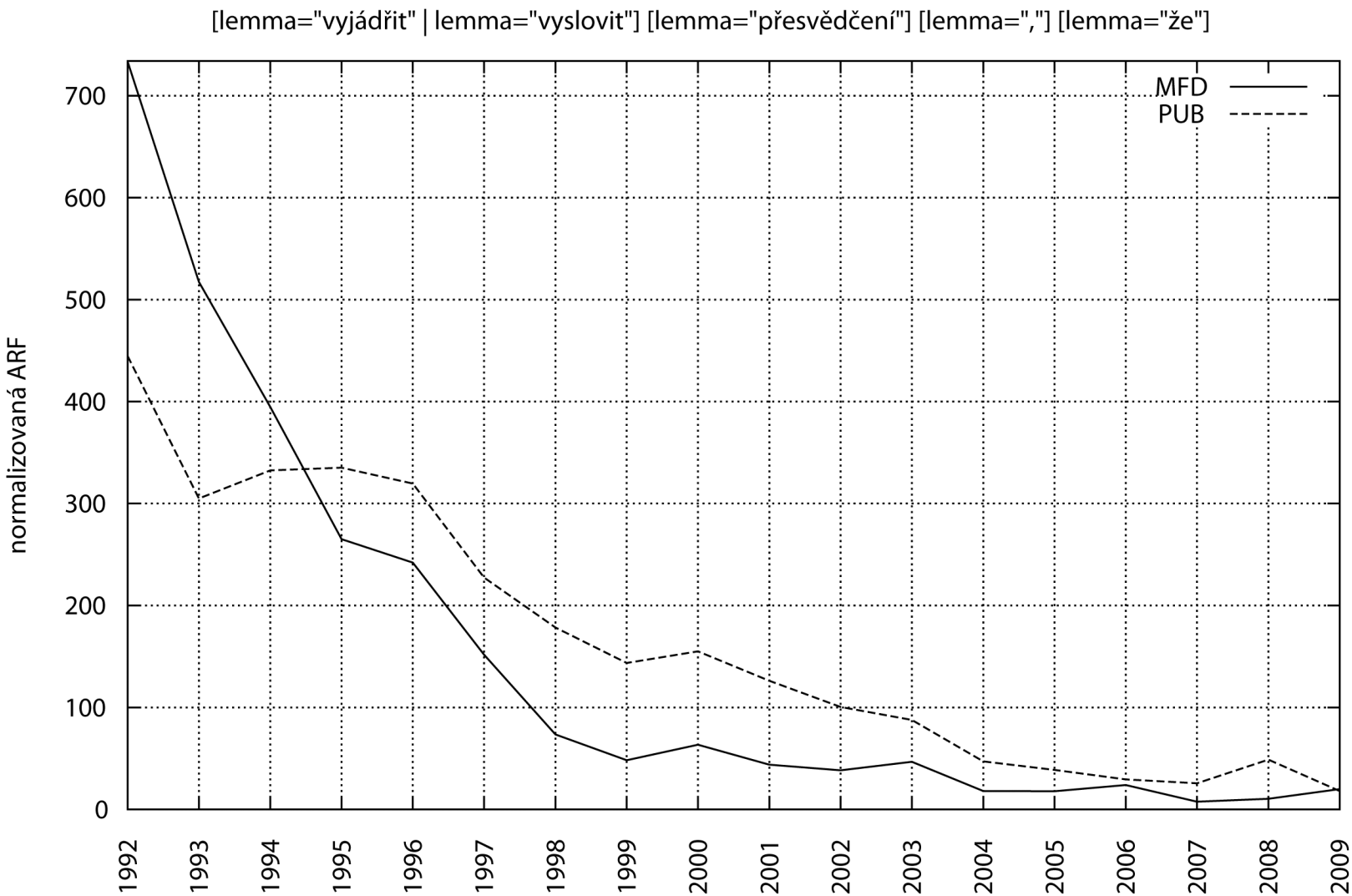
Skupina B:
fax, ing, marka, úloha, výbor, tisk, tajemník, eventuální, západoevropský; sdělovací prostředky, cenný papír, liberalizace cen, pluralitní demokracie, národní majetek...

Trend klesající

[lemma="volba" & tag="...P.*"]
normalizovaná ARF
20000
15000
10000
5000
0
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
MFD
PUB


[lemma="volba" & tag="...S.*"]
normalizovaná ARF
4500
4000
3500
3000
2500
2000
1500
1000
500
0
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
MFD
PUB


[lemma="lídr"]
normalizovaná ARF
0
500
1000
1500
2000
2500
3000
3500
4000
4500
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
MFD
PUB


[lemma="vyjádřit" | lemma="vyslovit"] [lemma="přesvědčení"] [lemma=","] [lemma="že"]
MFD
PUB
normalizovaná ARF
0
100
200
300
400
500
600
700
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009

Budování korpusů

Typy korpusů

- reprezentativní (vyvážené) $\times$ oportunistické

Typy korpusů

- reprezentativní (vyvážené) × oportunistické
- referenční × otevřené

Typy korpusů

- reprezentativní (vyvážené) × oportunistické
- referenční × otevřené
- mluvený jazyk × psaný jazyk × obojí

Typy korpusů

- reprezentativní (vyvážené) × oportunistické
- referenční × otevřené
- mluvený jazyk × psaný jazyk × obojí
- synchronní × diachronní

Typy korpusů

- reprezentativní (vyvážené) × oportunistické
- referenční × otevřené
- mluvený jazyk × psaný jazyk × obojí
- synchronní × diachronní
- jednojazyčné × vícejazyčné (paralelní)

Typy korpusů

- reprezentativní (vyvážené) × oportunistické
- referenční × otevřené
- mluvený jazyk × psaný jazyk × obojí
- synchronní × diachronní
- jednojazyčné × vícejazyčné (paralelní)
- obecné × specializované

Typy korpusů

- reprezentativní (vyvážené) × oportunistické
- referenční × otevřené
- mluvený jazyk × psaný jazyk × obojí
- synchronní × diachronní
- jednojazyčné × vícejazyčné (paralelní)
- obecné × specializované

Korpusy ve světě: BNC, CoCA, DeReKo, SNK...

Vytváření korpusů

Fáze přípravy korpusu:

1. sběr textů

Vytváření korpusů

Fáze přípravy korpusu:

1. sběr textů
2. konverze a čištění

Vytváření korpusů

Fáze přípravy korpusu:

1. sběr textů
2. konverze a čištění
3. přidělení lingvistické (strukturní) anotace

Vytváření korpusů

Fáze přípravy korpusu:

1. sběr textů
2. konverze a čištění
3. přidělení lingvistické (strukturní) anotace
4. vyvažování proporcí

Vytváření korpusů

Fáze přípravy korpusu:

1. sběr textů
2. konverze a čištění
3. přidělení lingvistické (strukturní) anotace
4. vyvažování proporcí
5. morfologie – dodání všech možných lemmat a tagů

Vytváření korpusů

Fáze přípravy korpusu:

1. sběr textů
2. konverze a čištění
3. přidělení lingvistické (strukturní) anotace
4. vyvažování proporcí
5. morfologie – dodání všech možných lemmat a tagů
6. disambiguace (zjednoznačnění) – stochastická, pravidlová

Reprezentativnost

Co má/může korpus reprezentovat?

1. produkci textů

Reprezentativnost

Co má/může korpus reprezentovat?

1. produkci textů
2. recepci textů

Reprezentativnost

Co má/může korpus reprezentovat?

1. produkci textů
2. recepci textů
3. texty/styly jako entity

Reprezentativnost

Co má/může korpus reprezentovat?

1. produkci textů
2. recepci textů
3. texty/styly jako entity

SYN2010:

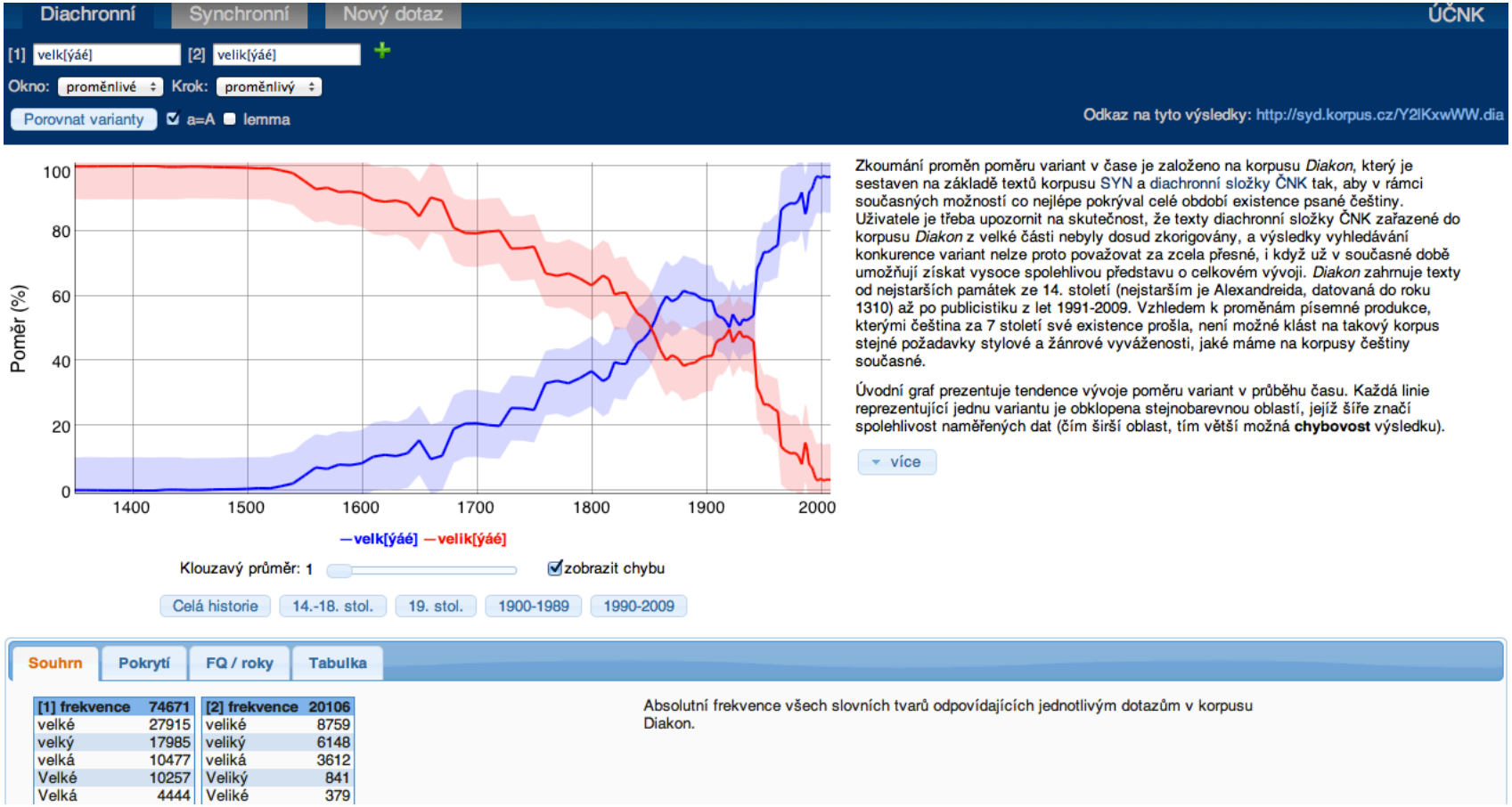
Poziční značky

word	lemma	tag
Po	po	RR--6----------
promoci	promoce	NNFS6-----A-----
na	na	RR--6----------
londýnské	londýnský	AAFS6----1A-----
universitě	univerzita	NNFS6-----A-----
odjel	odjet	VpMS---1R-AA---P
jsem	být	VB-S---1P-AA---I
roku	rok	NNIS2-----A---1-
1878	1878	C=-------------
do	do	RR--2----------
Netley	Netley	X@-------------
na	na	RR--4----------
školení	školení	NNNS4-----A-----
vojenských	vojenský	AAMP2----1A-----
chirurgů	chirurg	NNMP2-----A-----
.	.	Z:-------------

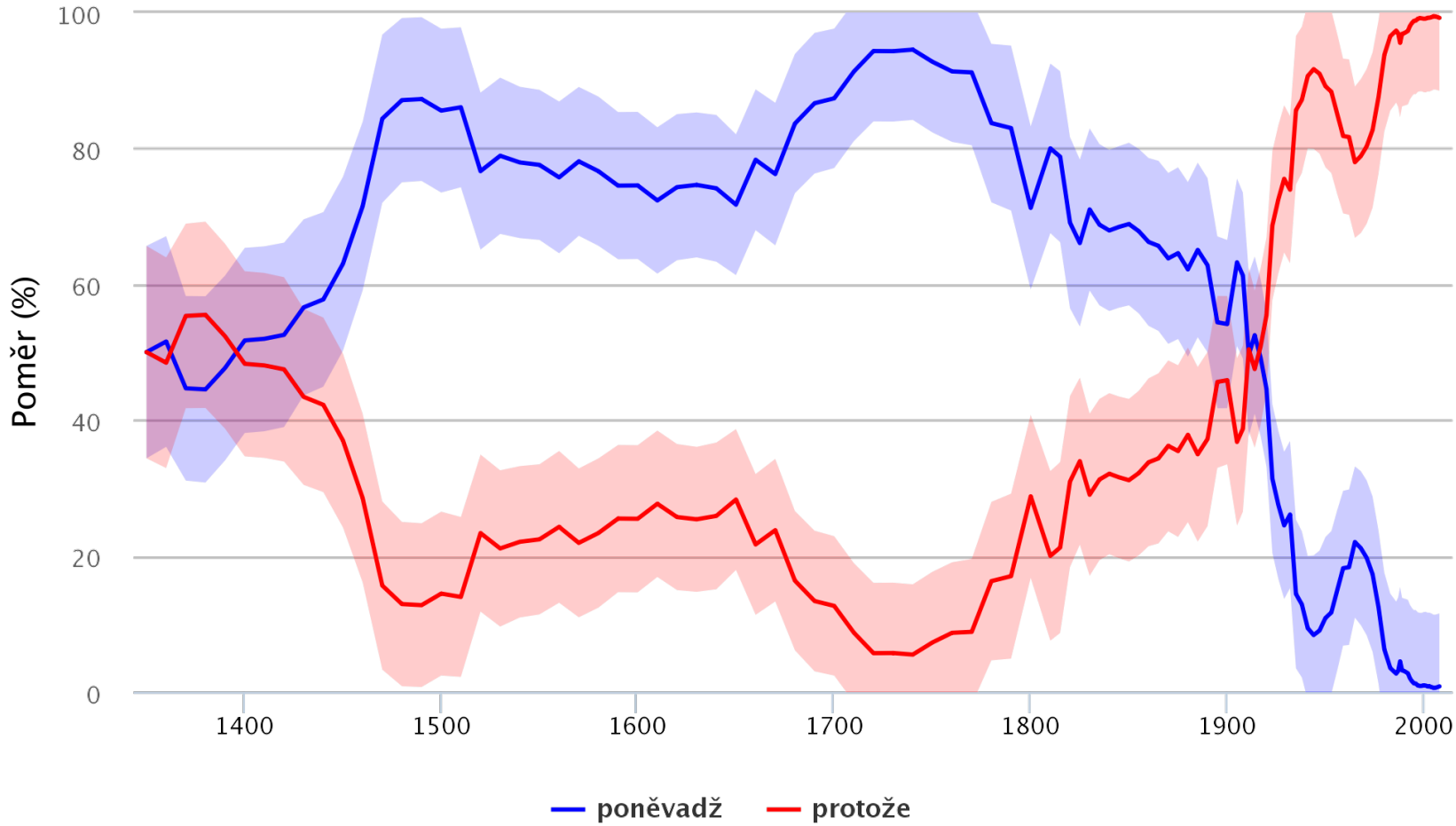
SyD – synchronní a diachronní analýza variant

http://syd.korpus.cz

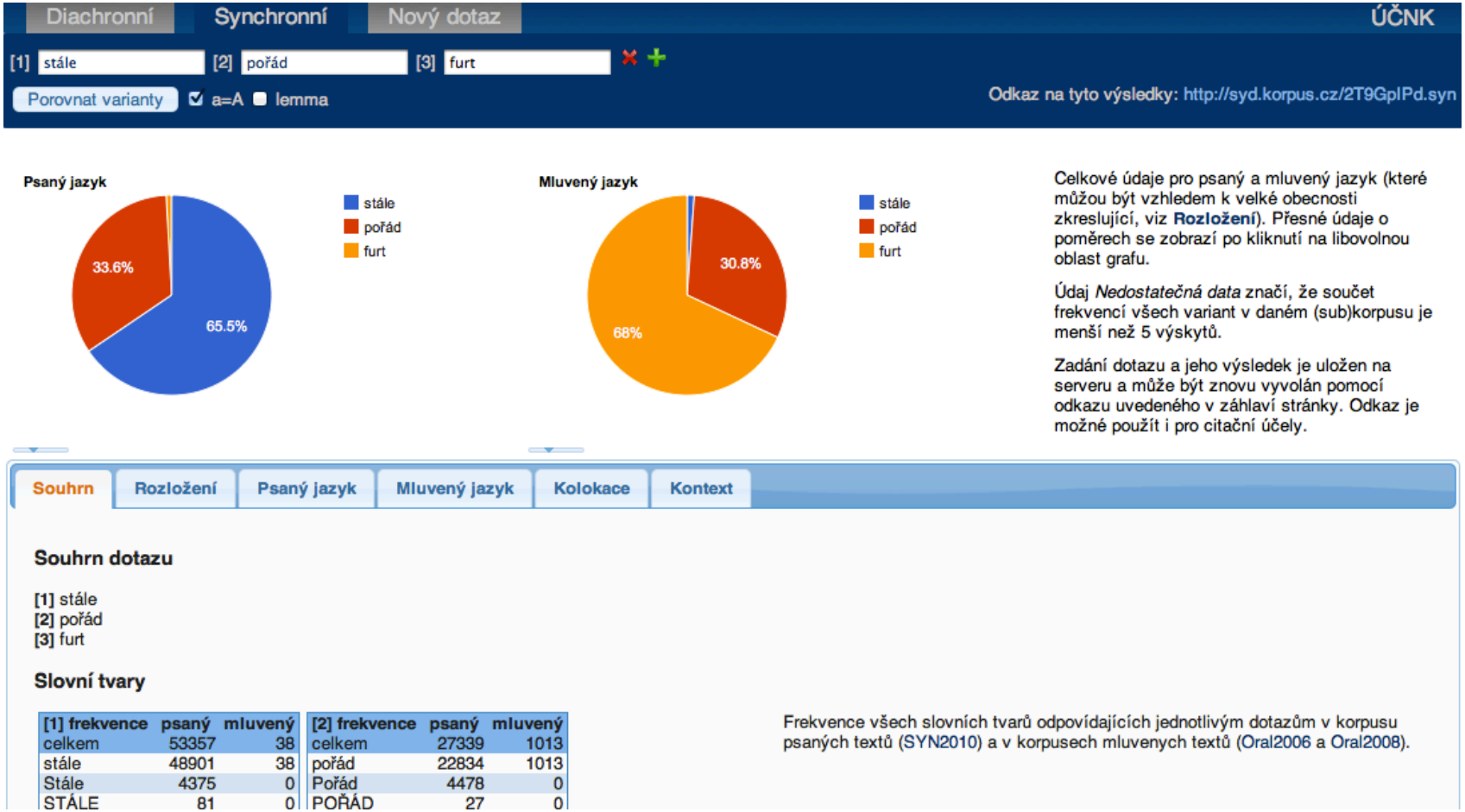
SyD – diachronní analýza variant



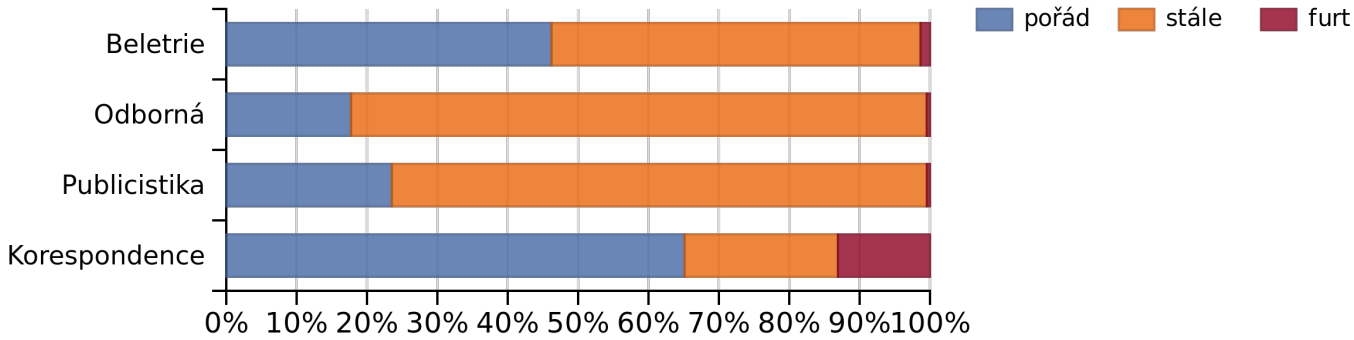
SyD – diachronní analýza variant



SyD – synchronní analýza variant



SyD – synchronní analýza variant



Využití korpusů

Využití korpusů

Korpus jako obecný zdroj informací o jazyce slouží

- lingvistickým účelům

Využití korpusů

Korpus jako obecný zdroj informací o jazyce slouží

- lingvistickým účelům
  - vytváření slovníků (in progress)

Využití korpusů

Korpus jako obecný zdroj informací o jazyce slouží

- lingvistickým účelům
  - vytváření slovníků (in progress)
  - vytváření gramatik (MSČ, AGSČ)

Využití korpusů

Korpus jako obecný zdroj informací o jazyce slouží

- lingvistickým účelům
  - vytváření slovníků (in progress)
  - vytváření gramatik (MSČ, AGSČ)
  - studium jazykové variability

Využití korpusů

Korpus jako obecný zdroj informací o jazyce slouží

- lingvistickým účelům
  - vytváření slovníků (in progress)
  - vytváření gramatik (MSČ, AGSČ)
  - studium jazykové variability
  - ...

Využití korpusů

Korpus jako obecný zdroj informací o jazyce slouží

- lingvistickým účelům
  - vytváření slovníků (in progress)
  - vytváření gramatik (MSČ, AGSČ)
  - studium jazykové variability
  - ...
- psychologickému výzkumu

Využití korpusů

Korpus jako obecný zdroj informací o jazyce slouží

- lingvistickým účelům
  - vytváření slovníků (in progress)
  - vytváření gramatik (MSČ, AGSČ)
  - studium jazykové variability
  - ...
- psychologickému výzkumu
  - frekvenční charakteristika stimulů

Využití korpusů

Korpus jako obecný zdroj informací o jazyce slouží

- lingvistickým účelům
  - vytváření slovníků (in progress)
  - vytváření gramatik (MSČ, AGSČ)
  - studium jazykové variability
  - ...
- psychologickému výzkumu
  - frekvenční charakteristika stimulů
  - asociační testy

Využití korpusů

Korpus jako obecný zdroj informací o jazyce slouží

- lingvistickým účelům
  - vytváření slovníků (in progress)
  - vytváření gramatik (MSČ, AGSČ)
  - studium jazykové variability
  - ...
- psychologickému výzkumu
  - frekvenční charakteristika stimulů
  - asociační testy
- historickému bádání

Využití korpusů

Korpus jako obecný zdroj informací o jazyce slouží

- lingvistickým účelům
  - vytváření slovníků (in progress)
  - vytváření gramatik (MSČ, AGSČ)
  - studium jazykové variability
  - ...
- psychologickému výzkumu
  - frekvenční charakteristika stimulů
  - asociační testy
- historickému bádání
  - analýza diskurzu (CADS)

Korpusová analýza diskurzu (CADS)

CADS = Corpus assisted discourse studies

- skupina metod využívajících k analýze diskurzu korpusy – nejužívanější je analýza klíčových slov (KWA)

Korpusová analýza diskurzu (CADS)

CADS = Corpus assisted discourse studies

- skupina metod využívajících k analýze diskurzu korpusy – nejužívanější je analýza klíčových slov (KWA)
- využívá rozsáhlé vzorky zkoumaného diskurzu i jednotlivé texty

Korpusová analýza diskurzu (CADS)

CADS = Corpus assisted discourse studies

- skupina metod využívajících k analýze diskurzu korpusy – nejužívanější je analýza klíčových slov (KWA)
- využívá rozsáhlé vzorky zkoumaného diskurzu i jednotlivé texty
- srovnání s referenčním korpusem (neutrální pozadí, reprezentace příjemce)

Korpusová analýza diskurzu (CADS)

CADS = Corpus assisted discourse studies

- skupina metod využívajících k analýze diskurzu korpusy – nejužívanější je analýza klíčových slov (KWA)
- využívá rozsáhlé vzorky zkoumaného diskurzu i jednotlivé texty
- srovnání s referenčním korpusem (neutrální pozadí, reprezentace příjemce)

Klíčová slova (keywords) = jednotky, které se ve zkoumaném textu/korpusu vyskytují se statisticky signifikantně vyšší frekvencí než v referenčním korpusu.

Analýza klíčových slov

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

Klíčová slova jsou východiskem pro další interpretaci

MD-CADS = proměny diskurzu v čase

Vhodná data pro MD-CADS

- prezidentské novoroční projevy – původně pouze Gustáv Husák (1975–1989), později i další

MD-CADS = proměny diskurzu v čase

Vhodná data pro MD-CADS

- prezidentské novoroční projevy – původně pouze Gustáv Husák (1975–1989), později i další
- považovány za nudné, monotonní, plné klišé – `www.youtube.com/watch?v=QiBm4YX9y24`

MD-CADS = proměny diskurzu v čase

Vhodná data pro MD-CADS

- prezidentské novoroční projevy – původně pouze Gustáv Husák (1975–1989), později i další
- považovány za nudné, monotonní, plné klišé – `www.youtube.com/watch?v=QiBm4YX9y24`
- optimální pro testování: stejné prostředí, situace, autor $\times$ liší se časem (a tématem)

MD-CADS = proměny diskurzu v čase

Vhodná data pro MD-CADS

- prezidentské novoroční projevy – původně pouze Gustáv Husák (1975–1989), později i další
- považovány za nudné, monotonní, plné klišé – `www.youtube.com/watch?v=QiBm4YX9y24`
- optimální pro testování: stejné prostředí, situace, autor $\times$ liší se časem (a tématem)
- referenční korpusy

MD-CADS = proměny diskurzu v čase

Vhodná data pro MD-CADS

- prezidentské novoroční projevy – původně pouze Gustáv Husák (1975–1989), později i další
- považovány za nudné, monotonní, plné klišé – `www.youtube.com/watch?v=QiBm4YX9y24`
- optimální pro testování: stejné prostředí, situace, autor $\times$ liší se časem (a tématem)
- referenční korpusy
  - SYN2010 – 100 mil. slov (2005–2009), reprezentuje současného čtenáře

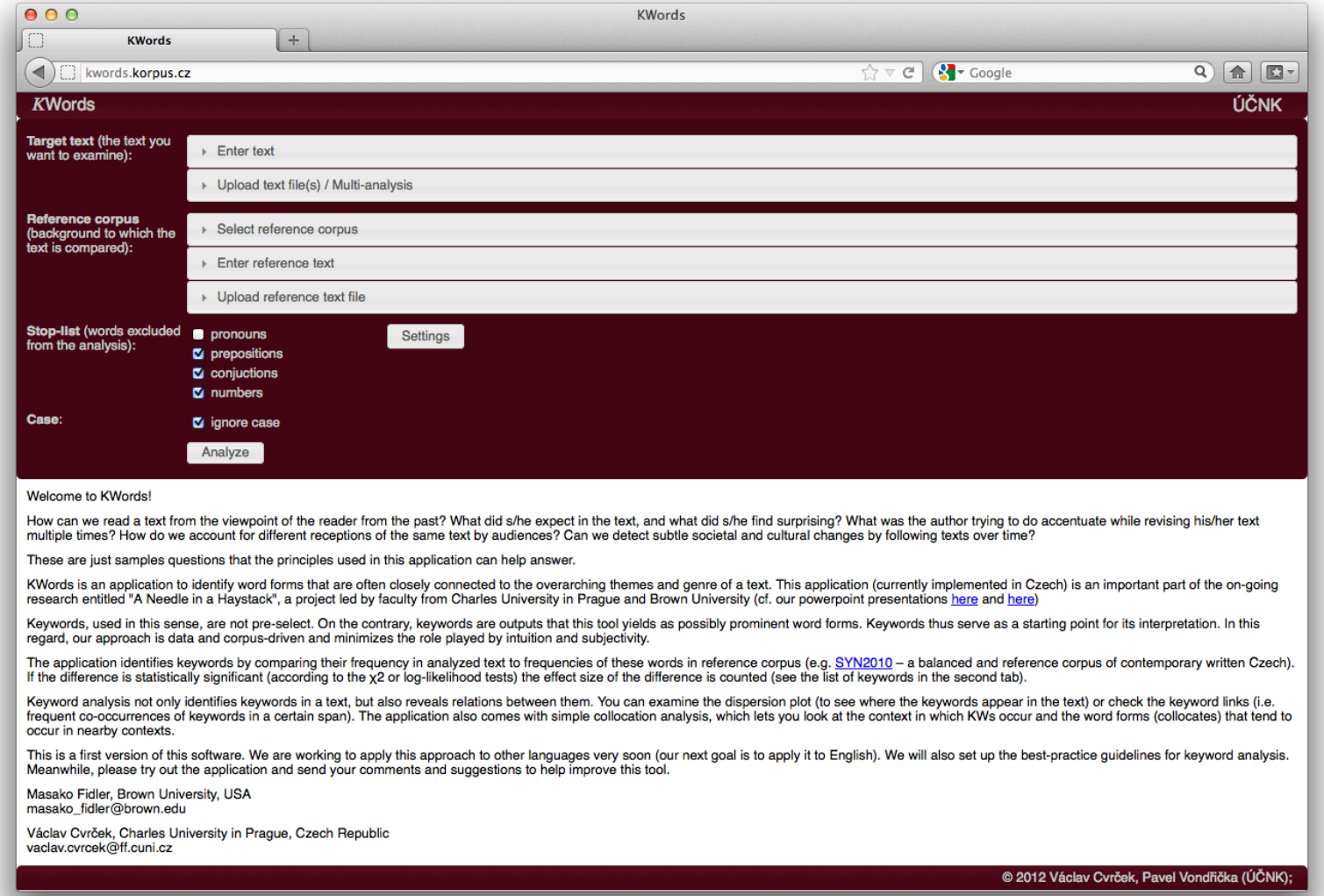
MD-CADS = proměny diskurzu v čase

Vhodná data pro MD-CADS

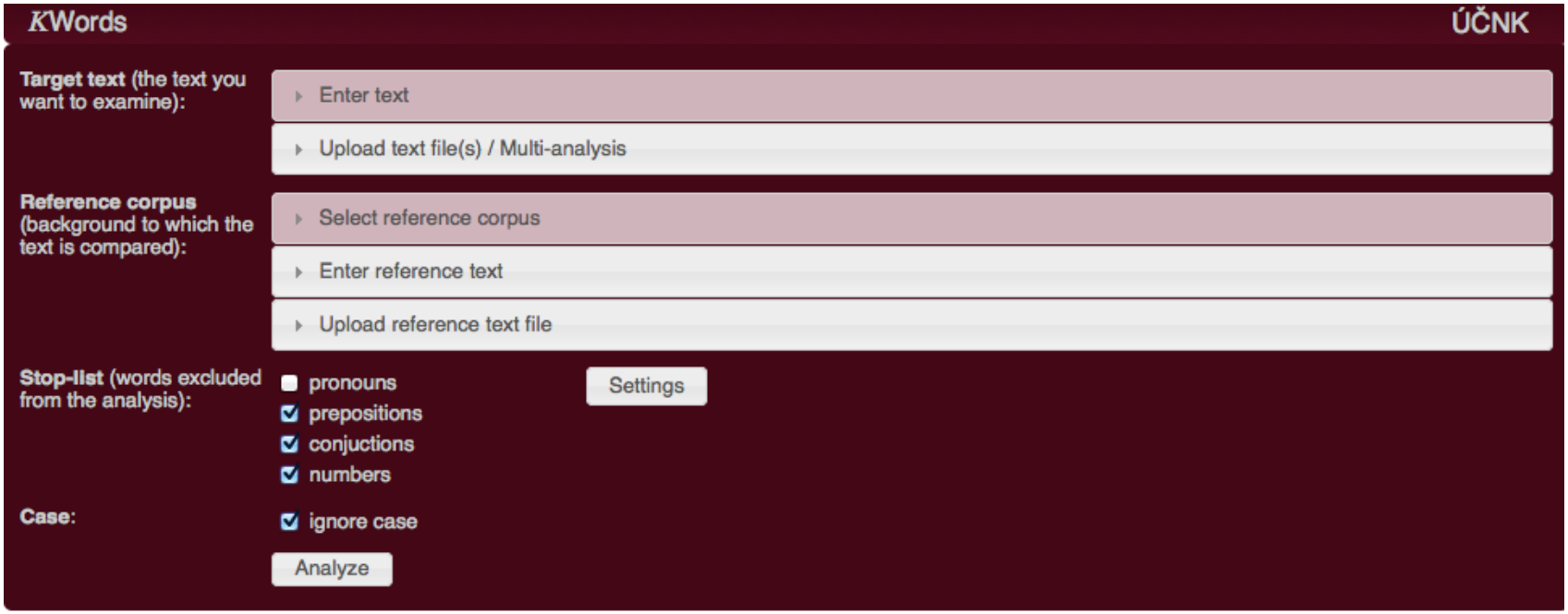
- prezidentské novoroční projevy – původně pouze Gustáv Husák (1975–1989), později i další
- považovány za nudné, monotonní, plné klišé – `www.youtube.com/watch?v=QiBm4YX9y24`
- optimální pro testování: stejné prostředí, situace, autor $\times$ liší se časem (a tématem)
- referenční korpusy
  - SYN2010 – 100 mil. slov (2005–2009), reprezentuje současného čtenáře
  - Totalita – 15 mil. slov (1952–1977) *Rudé právo* a další dobové tiskoviny $\sim$ čtenář v minulosti

Aplikace KWords

http://kwords.korpus.cz

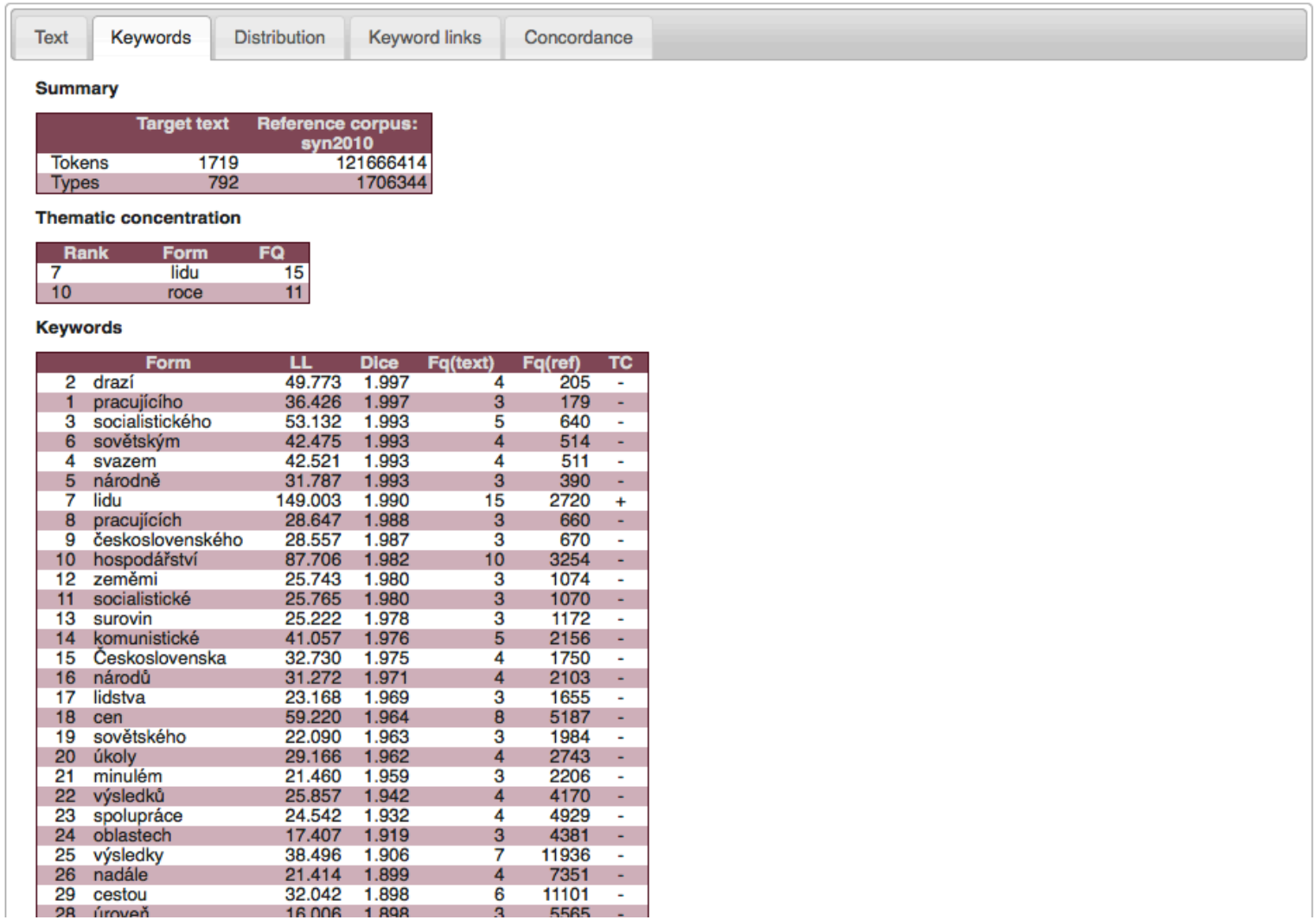


KWords: vkládání analyzovaného textu



KWords: seznam klíčových slov

2 typy prominentních jednotek: klíčová slova and tématická koncentrace



Text | Keywords | Distribution | Keyword links | Concordance

Summary

	Target text	Reference corpus: syn2010
Tokens	1719	121666414
Types	792	1706344

Thematic concentration

Rank	Form	FQ
7	lidu	15
10	roce	11

Keywords

	Form	LL	Dice	Fq(text)	Fq(ref)	TC
2	drazí	49.773	1.997	4	205	-
1	pracujícího	36.426	1.997	3	179	-
3	socialistického	53.132	1.993	5	640	-
6	sovětským	42.475	1.993	4	514	-
4	svazem	42.521	1.993	4	511	-
5	národně	31.787	1.993	3	390	-
7	lidu	149.003	1.990	15	2720	+
8	pracujících	28.647	1.988	3	660	-
9	československého	28.557	1.987	3	670	-
10	hospodářství	87.706	1.982	10	3254	-
12	zeměmi	25.743	1.980	3	1074	-
11	socialistické	25.765	1.980	3	1070	-
13	surovin	25.222	1.978	3	1172	-
14	komunistické	41.057	1.976	5	2156	-
15	Československa	32.730	1.975	4	1750	-
16	národů	31.272	1.971	4	2103	-
17	lidstva	23.168	1.969	3	1655	-
18	cen	59.220	1.964	8	5187	-
19	sovětského	22.090	1.963	3	1984	-
20	úkoly	29.166	1.962	4	2743	-
21	minulém	21.460	1.959	3	2206	-
22	výsledků	25.857	1.942	4	4170	-
23	spolupráce	24.542	1.932	4	4929	-
24	oblastech	17.407	1.919	3	4381	-
25	výsledky	38.496	1.906	7	11936	-
26	nadále	21.414	1.899	4	7351	-
29	cestou	32.042	1.898	6	11101	-
28	úroveň	16.006	1.898	3	5565	-

Keyword links

Links = vazby mezi klíčovými slovy

Vzdálené KW Links = klíčová slova objevující se ve vzdáleném kontextu (-15;-5) a (5;15); indikují blízkost témat

Keyword links

Links = vazby mezi klíčovými slovy

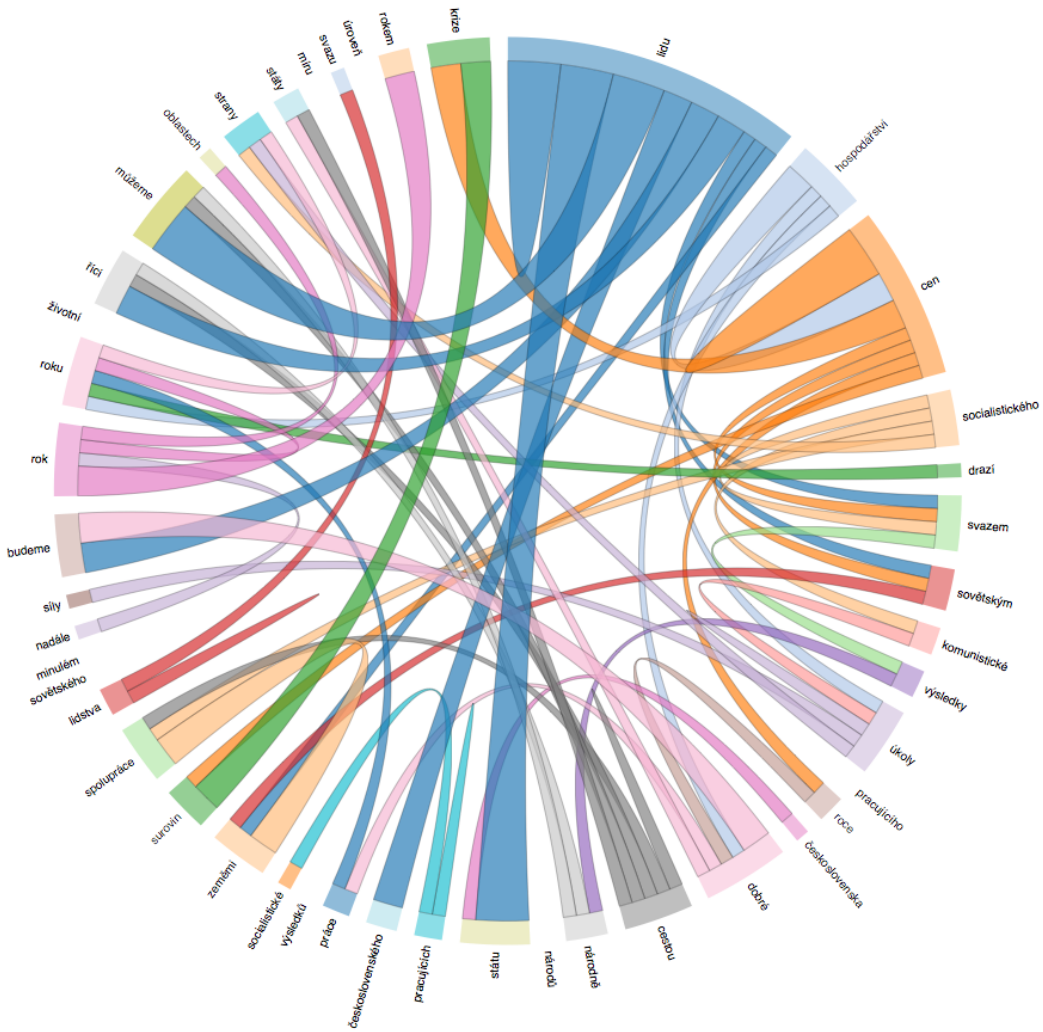
Vzdálené KW Links = klíčová slova objevující se ve vzdáleném kontextu (-15;-5) a (5;15); indikují blízkost témat

Bezprostřední KW Links = souvýskyt dvou nebo více klíčových slov v blízkém kontextu (-2;2); prominentní víceslovné jednotky (např. *národní fronta*)

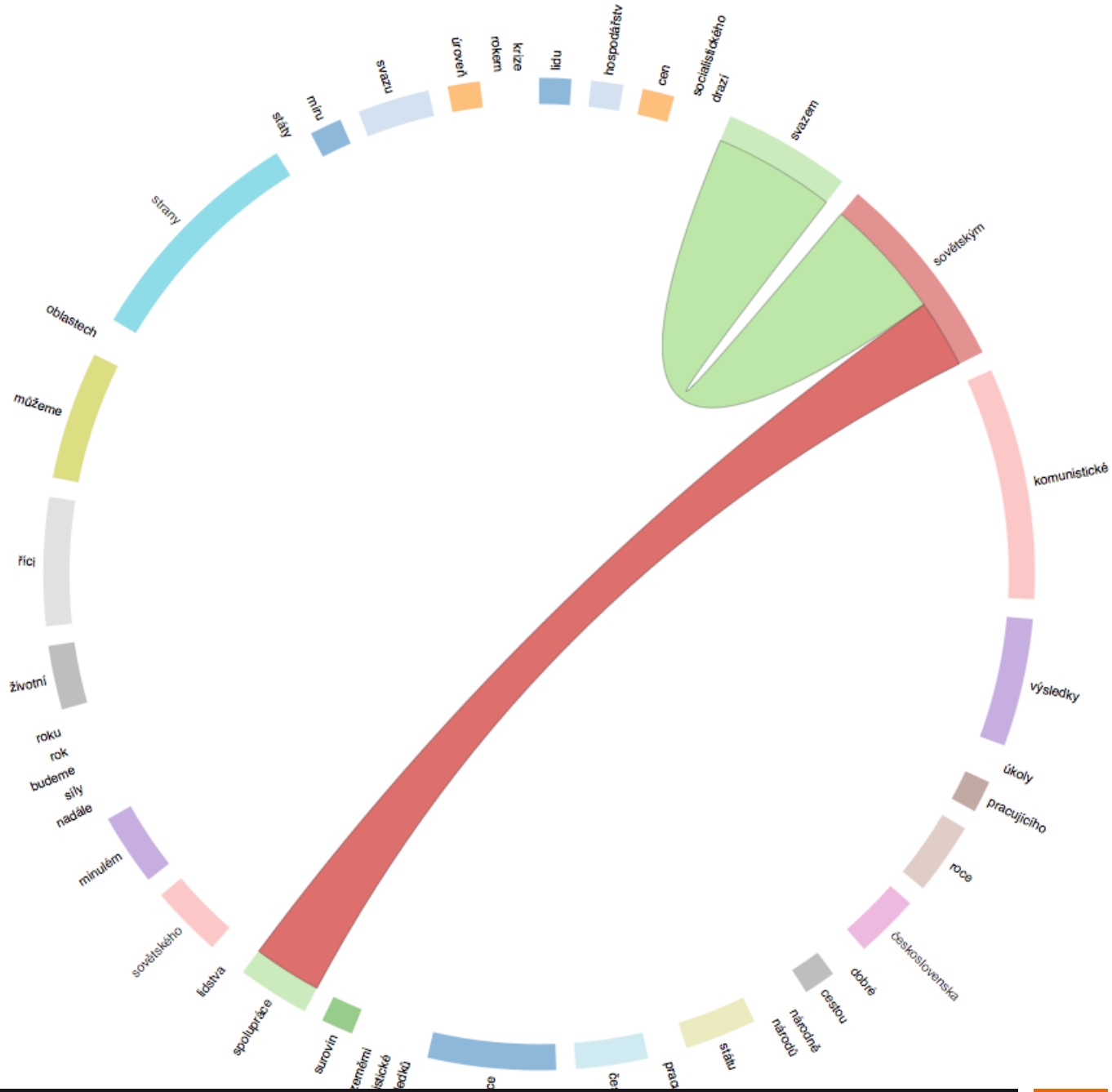
Text | Keywords | Distribution | **Keyword links** | Concordance

Presets: distant KW links

☑ hide single links



Save graph as PDF

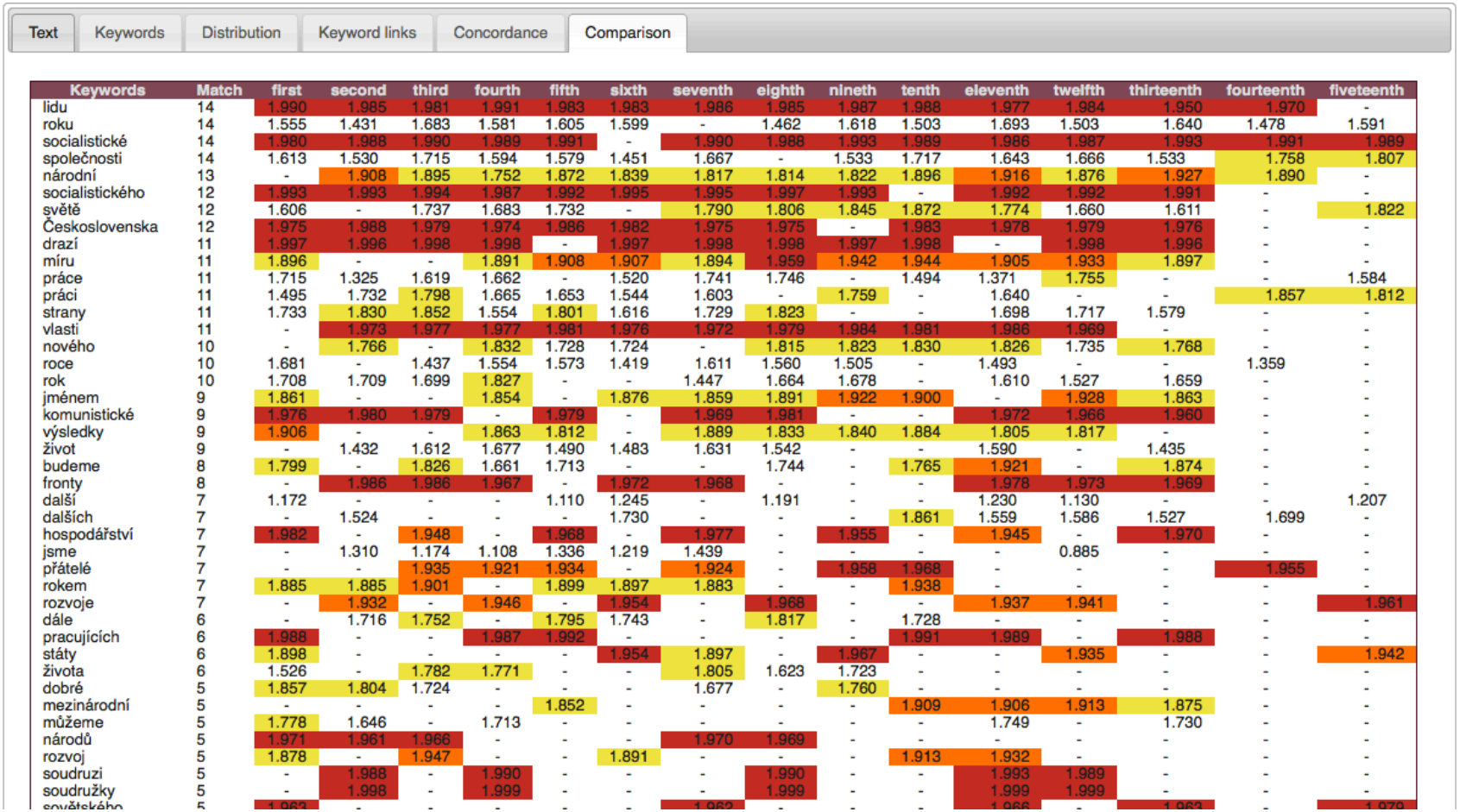

míru
svazu
úroveň
rokem
krize
lidu
hospodářství
cen
socialistického
drazí
svazem
sovětským
komunistické
výsledky
úkoly
pracujícího
roce
československa
dobré
cestou
národně
národů
státu
spolupráce
surovin
sovětského
lidstva
minulém
nadále
síly
budeme
rok
roku
životní
řící
můžeme
oblastech
strany
státy

KWords: konkordance

Text | Keywords | Distribution | Keyword links | Concordance

	most frequent words on preceding position	keyword	most frequent words on following position
	prací (1) • zápasu (1) • 1975 (1) • láskou (1)	budeme (5)	si (1) • ze (1) • vzpomínat (1) • udržovat (1)
	světových (2) • zvýšených (2) • zvýšení (1) • spotřebitelských (1)	cen (8)	a (2) • surovin (2) • , (1) • těchto (1)
	šel (1) • prospěšnou (1) • , (1) • ale (1)	cestou (6)	mírové (1) • využití (1) • míru (1) • naše (1)
	strany (2) • okupaci (1) • socialistického (1)	československa (4)	, (3) • . (1)
	a (1) • , (1) • obnovu (1)	československého (3)	státu (3)
	na (2) • tyto (1) • existují (1) • 1975 (1)	dobré (7)	výsledky (1) • vztahy (1) • pozitivní (1) • předpoklady (1)
	. (2) • , (2)	drazí (4)	spoluobčané (2) • přátelé (2)
	naše (3) • našeho (2) • úkoly (1) • rozvoj (1)	hospodářství (10)	, (3) • má (1) • mohlo (1) • opírá (1)
[v] POS	[ ] LEFT CONTEXT		RIGHT CONTEXT [ ]
122	přiřadil k několika předcházejícím rokům v rozvoji našeho	hospodářství	, životní úrovně a celé společnosti . dobře
355	a považuje je za samozřejmé . ale rozvoj	hospodářství	a celé společnosti vždy přináší i problémy ,
411	cen těchto produktů . i když se naše	hospodářství	opírá především o spolupráci se sovětským svazem a
441	nás nemalý . s pohybem cen ve světovém	hospodářství	je třeba počítat i v roce 1975 a
554	cestu překonávání vlivu růstu světových cen na naše	hospodářství	aktivními opatřeními ve výrobě zvyšováním její efektivnosti ,
619	se ukázala přednost socialistického zřízení , jeho plánovitého	hospodářství	i politiky komunistické strany , která na první
1178	se všemi státy . máme co nabídnout v	hospodářství	, ve vědě , technice , kultuře i
1342	jejich hlavním smyslem je zajistit další rozvoj našeho	hospodářství	, životní úroveň a šťastný život našeho lidu
1457	velké praxe nás však přesvědčuje , že naše	hospodářství	má ještě mnoho rezerv a je schopno dosáhnout
1551	tyto úkoly splnit . v souladu s úkoly	hospodářství	mohlo listopadové plenární zasedání komunistické strany z minulého
	politiky (2) • sjedu (1) • zasedání (1) • výboru (1)	komunistické (5)	strany (5)
	světě (1) • morální (1) • energetické (1)	krize (3)	buržoazního (1) • energetiky (1) • na (1)
	zničení (1) • prospěch (1) • zkáze (1)	lidstva (3)	. (2) • vůbec (1)
	našeho (9) • pracujícího (2) • jejího (1) • jeho (1)	lidu (15)	. (4) • , (4) • a (2) • i (2)
	v (3)	minulém (3)	roce (3)
	cestou (1) • či (1) • v (1)	míru (3)	, (2) • pro (1)
	. (2) • že (1) • úspěchy (1) • let (1)	můžeme (5)	říci (4) • být (1)
	i (4)	nadále (4)	dařila (1) • zabezpečeny (1) • napomáhat (1) • udržet (1)
	odkazu (1) • vyvrcholení (1) • na (1)	národně (3)	osvobozeneckého (2) • osvobozenecký (1)
	našich (2) • jiných (1) • bratrských (1)	národů (4)	a (2) • , (1) • při (1)
	všech (1) • rozsáhlých (1) • ostatních (1)	oblastech (3)	pro (1) • naší (1) • . (1)
	výsledky (4) • naše (1) • naší (1) • iniciativní (1)	práce (8)	našich (1) • našeho (1) • miliónů (1) • třiceti (1)
	všeho (1) • jednoty (1) • blaho (1)	pracujícího (3)	lidu (2) • člověka (1)
	nejistotu (1) • bedra (1) • našich (1)	pracujících (3)	. (2) • mohli (1)
	v (5) • minulém (3) • tomto (1) • novém (1)	roce (11)	1975 (2) • 1974 (2) • , (1) • pozitivních (1)

KWords: simultánní analýza více textů



Text | Keywords | Distribution | Keyword links | Concordance | Comparison

Keywords	Match	first	second	third	fourth	fifth	sixth	seventh	eighth	nineth	tenth	eleventh	twelfth	thirteenth	fourteenth	fiveteenth
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	-	1.533	1.717	1.643	1.666	1.533	1.758	1.807
národní	13	-	1.908	1.895	1.752	1.872	1.839	1.817	1.814	1.822	1.896	1.916	1.876	1.927	1.890	-
socialistického	12	1.993	1.993	1.994	1.987	1.992	1.995	1.995	1.997	1.993	-	1.992	1.992	1.991	-	-
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	1.986	1.982	1.975	1.975	-	1.983	1.978	1.979	1.976	-	-
drazí	11	1.997	1.996	1.998	1.998	-	1.997	1.998	1.998	1.997	1.998	-	1.998	1.996	-	-
míru	11	1.896	-	-	1.891	1.908	1.907	1.894	1.959	1.942	1.944	1.905	1.933	1.897	-	-
práce	11	1.715	1.325	1.619	1.662	-	1.520	1.741	1.746	-	1.494	1.371	1.755	-	-	1.584
práci	11	1.495	1.732	1.798	1.665	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	1.981	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	1.708	1.709	1.699	1.827	-	-	1.447	1.664	1.678	-	1.610	1.527	1.659	-	-
jménem	9	1.861	-	-	1.854	-	1.876	1.859	1.891	1.922	1.900	-	1.928	1.863	-	-
komunistické	9	1.976	1.980	1.979	-	1.979	-	1.969	1.981	-	-	1.972	1.966	1.960	-	-
výsledky	9	1.906	-	-	1.863	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	-	1.986	1.986	1.967	-	1.972	1.968	-	-	-	1.978	1.973	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	1.982	-	1.948	-	1.968	-	1.977	-	1.955	-	1.945	-	1.970	-	-
jsme	7	-	1.310	1.174	1.108	1.336	1.219	1.439	-	-	-	-	0.885	-	-	-
přátelé	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	-	1.954	-	1.968	-	-	1.937	1.941	-	-	1.961
dále	6	-	1.716	1.752	-	1.795	1.743	-	1.817	-	1.728	-	-	-	-	-
pracujících	6	1.988	-	-	1.987	1.992	-	-	-	-	1.991	1.989	-	1.988	-	-
státy	6	1.898	-	-	-	-	1.954	1.897	-	1.967	-	-	1.935	-	-	1.942
života	6	1.526	-	1.782	1.771	-	-	1.805	1.623	1.723	-	-	-	-	-	-
dobré	5	1.857	1.804	1.724	-	-	-	1.677	-	1.760	-	-	-	-	-	-
mezinárodní	5	-	-	-	-	1.852	-	-	-	-	1.909	1.906	1.913	1.875	-	-
můžeme	5	1.778	1.646	-	1.713	-	-	-	-	-	-	1.749	-	1.730	-	-
národů	5	1.971	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	-	-	-
soudružky	5	-	1.998	-	1.999	-	-	-	1.999	-	-	1.999	1.999	-	-	-
sovětského	5	1.963	-	-	-	-	-	1.962	-	-	-	1.966	-	1.963	-	1.979

Vliv referenčního korpusu

Jak dopadá srovnání podle různých referenčních korpusů?

- čím větší korpus, tím víc klíčových slov

Vliv referenčního korpusu

Jak dopadá srovnání podle různých referenčních korpusů?

- čím větší korpus, tím víc klíčových slov
- inventář klíčových slov se povětšinou neliší

Vliv referenčního korpusu

Jak dopadá srovnání podle různých referenčních korpusů?

- čím větší korpus, tím víc klíčových slov
- inventář klíčových slov se povětšinou neliší
- rozdíl je v míře prominence klíčových slov (pořadí)

Vliv referenčního korpusu

Jak dopadá srovnání podle různých referenčních korpusů?

- čím větší korpus, tím víc klíčových slov
- inventář klíčových slov se povětšinou neliší
- rozdíl je v míře prominence klíčových slov (pořadí)

Historický čtenář (Totalita)

→ Rozdíl v žánru

- Modální slovesa: *chci, mohu*
- Slovesa: 1. sg./pl.

Vliv referenčního korpusu

Jak dopadá srovnání podle různých referenčních korpusů?

- čím větší korpus, tím víc klíčových slov
- inventář klíčových slov se povětšinou neliší
- rozdíl je v míře prominence klíčových slov (pořadí)

Historický čtenář (Totalita)

→ Rozdíl v žánru

- Modální slovesa: *chci, mohu*
- Slovesa: 1. sg./pl.

Současný čtenář (SYN2010)

→ Historické reálie

- ideologie
- archaismy, historismy

KW Links: Klíčové slovo *komunistické* / *KSČ*

Rok	Tvar	Fq	Levý kontext*
1975	*komunistické*	5	jménem, orgány strany, “správná” politika
1976	*komunistické*	6	jménem, orgány strany, úspěšná politika
1977	*komunistické*	5	jménem, orgány strany, podpora politice
1978	–	–	–
1979	*komunistické*	5	jménem, orgány strany, podpora politice
1980	–	–	–
1981	*komunistické*	4	jménem, orgány strany, (sebe)kritika
1982	*komunistické*	5	jménem, orgány strany, podpora principů
1983	*KSČ*	5	jménem, orgány strany, “správná” politika
1984	–	–	–
1985	*komunistické*	4	jménem, orgány strany, důvěra v politiku
1986	*komunistické*	3	jménem, orgány strany
1987	*komunistické*	3	jménem, orgány strany, KSSS
1988	*KSČ*	3	orgány strany
1989	*KSČ*	4	orgány strany

*V pravém kontextu se vyskytuje vždy slovo *strany*.

Tématická výstavba a řečové strategie

- G. Husák (1975–1989) – 1975 za prezidenta Svobodu
- V. Havel (1990–2003) – 1993 prezidentský slib
- V. Klaus (2004–2011)
- M. Zeman (2013) – inaugurační slib a vánoční poselství

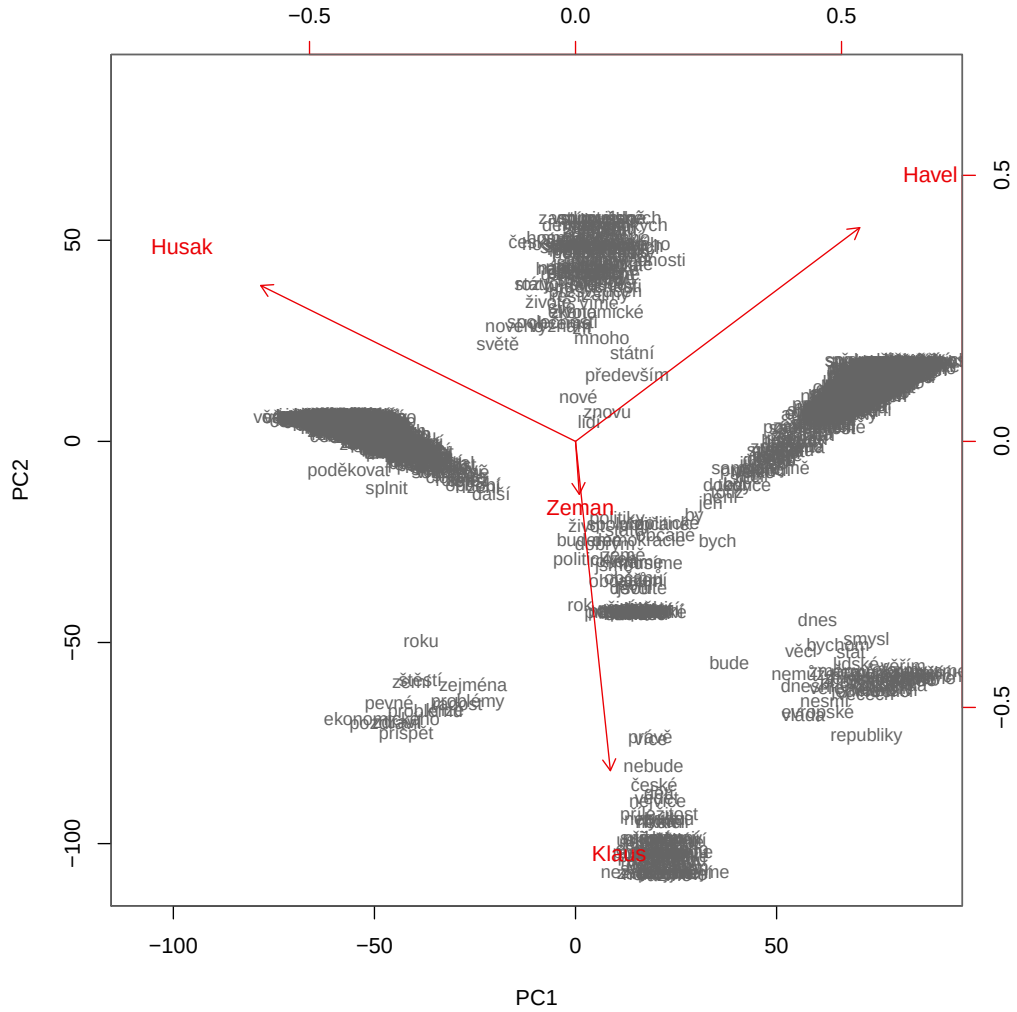
Tématická výstavba a řečové strategie

- G. Husák (1975–1989) – 1975 za prezidenta Svobodu
- V. Havel (1990–2003) – 1993 prezidentský slib
- V. Klaus (2004–2011)
- M. Zeman (2013) – inaugurační slib a vánoční poselství

	Projevy	Délka	Počet typů	Průměrná délka
Husák	15	22 058	4 162	1 470,5
Havel	14	33 153	7 457	2 368,1
Klaus	8	8 107	2 594	1 013,4
Zeman	2	1 896	894	948,0

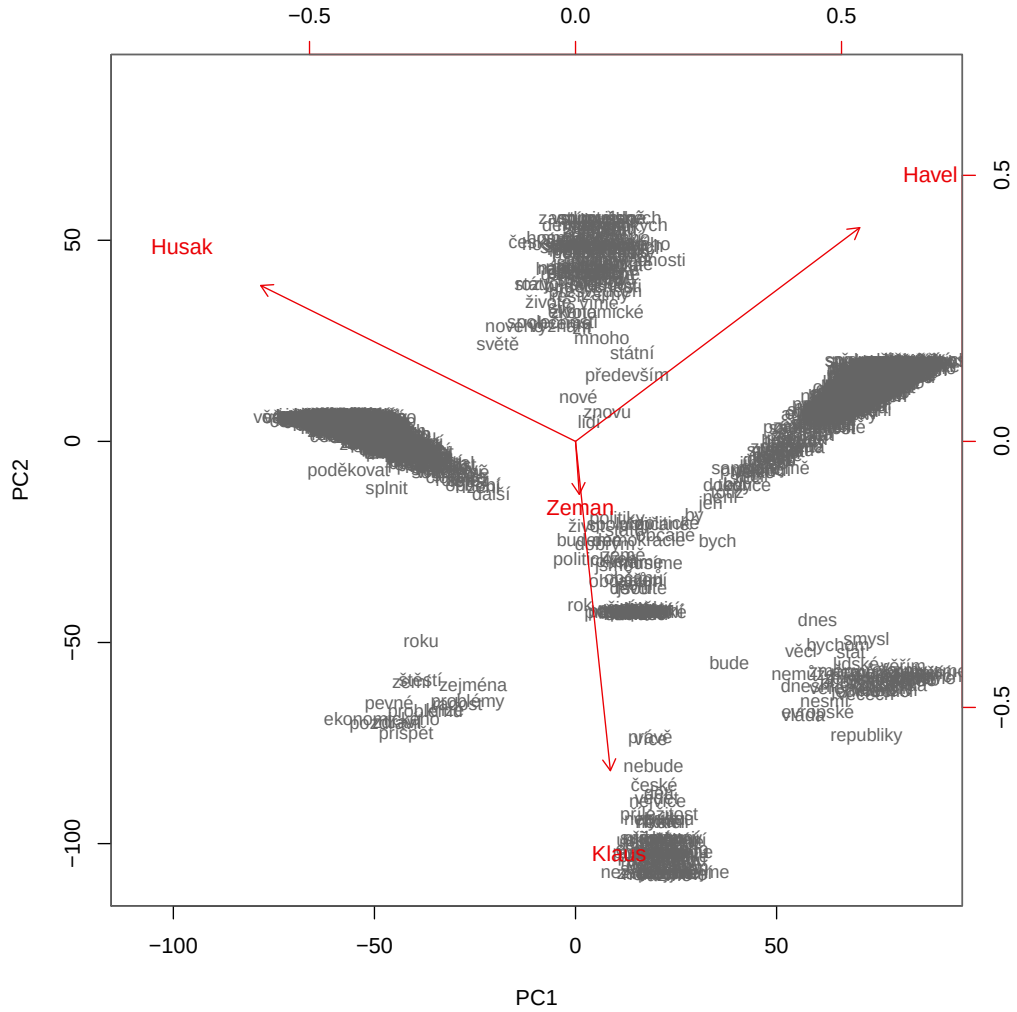
Tématická výstavba

- KWA textů podle mluvčích



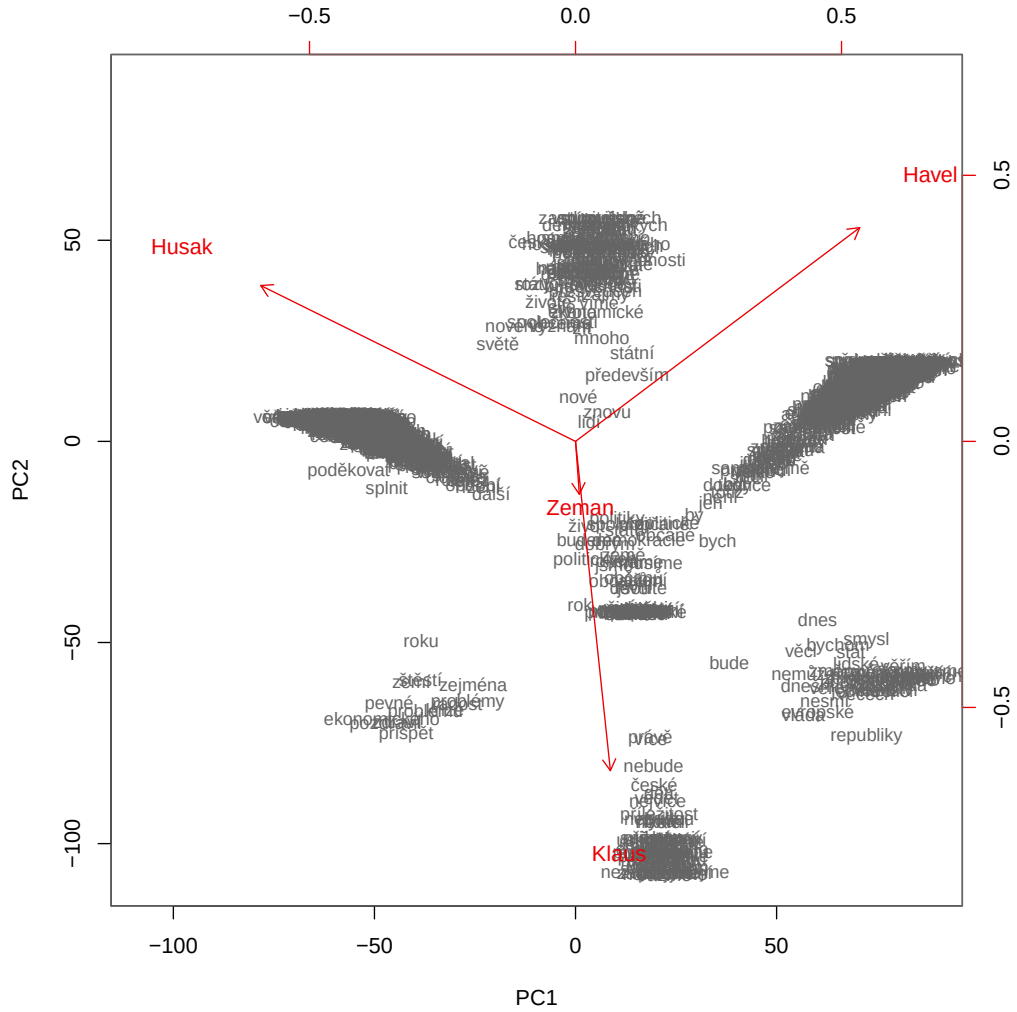
Tématická výstavba

- KWA textů podle mluvčích
- velice málo společných KWs



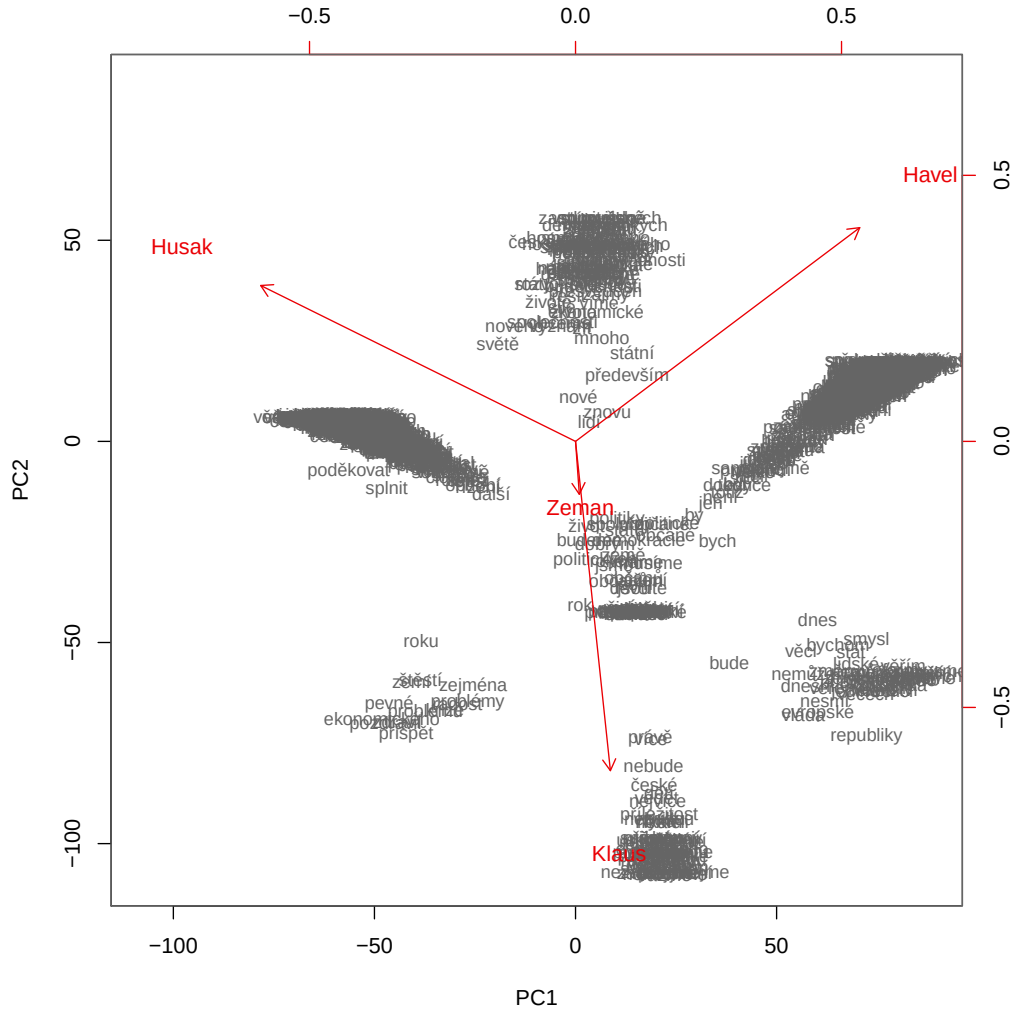
Tématická výstavba

- KWA textů podle mluvčích
- velice málo společných KWs
- obtížné tématické srovnání



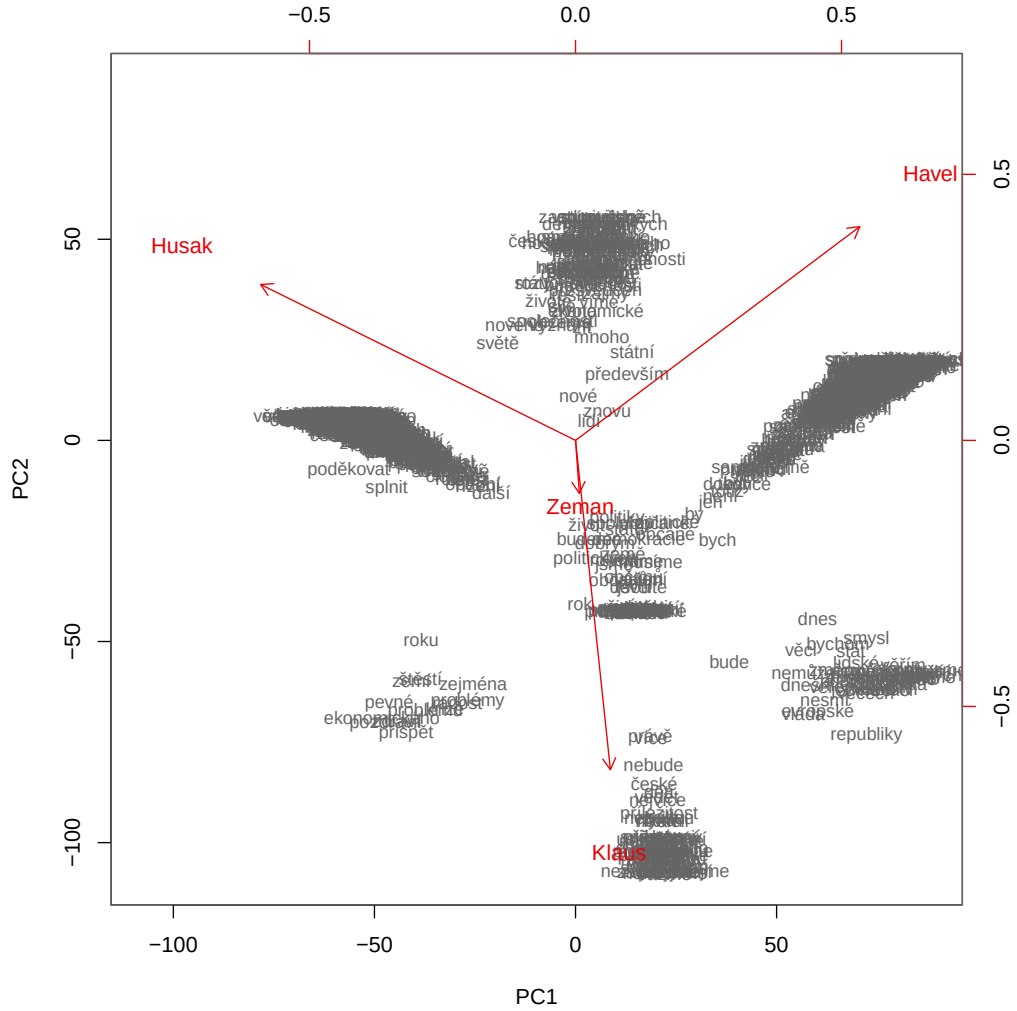
Tématická výstavba

- KWA textů podle mluvčích
- velice málo společných KWs
- obtížné tématické srovnání
- každý prezident má svoje specifická témata (daná agendou příslušného roku)



Tématická výstavba

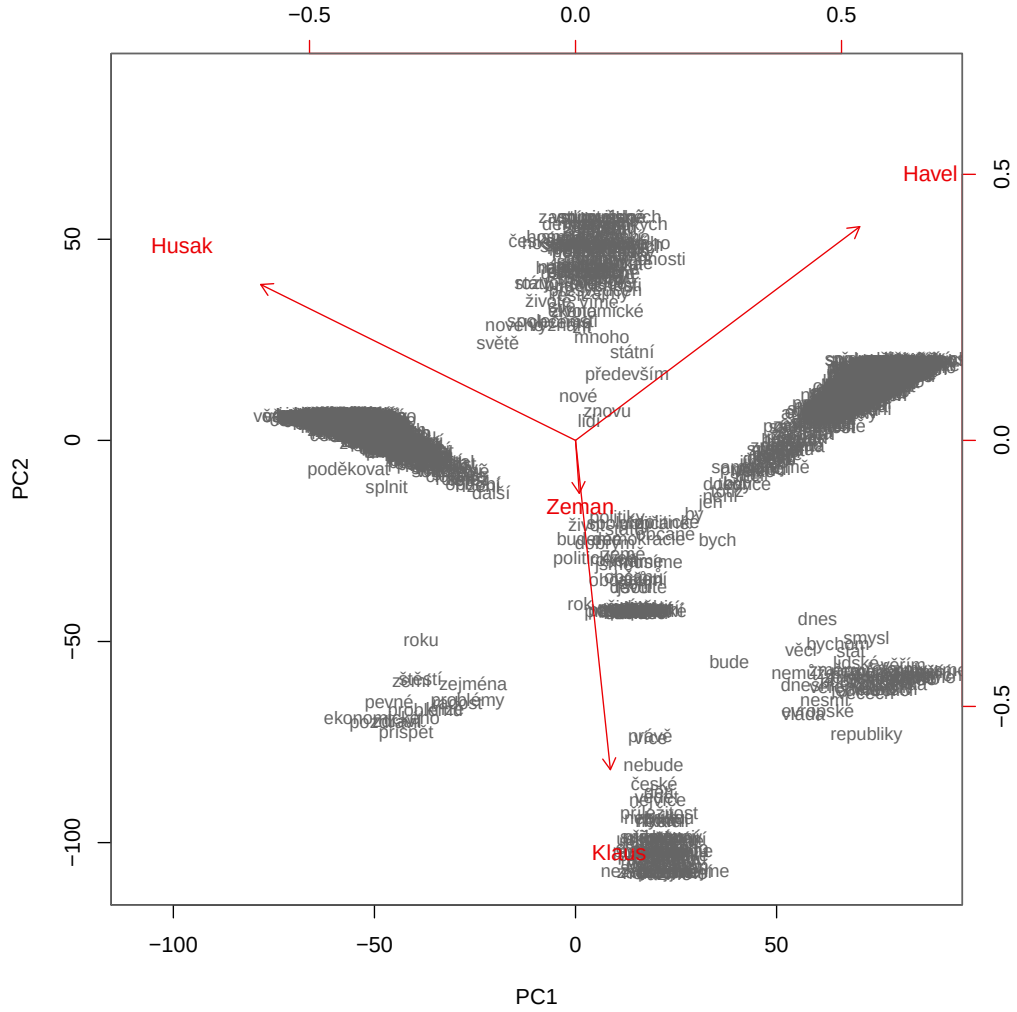
- KWA textů podle mluvčích
- velice málo společných KWs
- obtížné tématické srovnání
- každý prezident má svoje specifická témata (daná agendou příslušného roku)
- soubory textů od jednotlivých mluvčích – příliš heterogenní skupina (agregát)



Tématická výstavba

- KWA textů podle mluvčích
- velice málo společných KWs
- obtížné tématické srovnání
- každý prezident má svoje specifická témata (daná agendou příslušného roku)
- soubory textů od jednotlivých mluvčích – příliš heterogenní skupina (agregát)
- nedostatek textů od MZ

⇒ analýza řečnických strategií



Řečnické strategie

Zaměření na gramatické a flektivní charakteristiky KWs

- při porovnávání prezidentů bereme v úvahu nejen frekvenci tvarů, ale také jejich prominenci (DIN)

Řečnické strategie

Zaměření na gramatické a flektivní charakteristiky KWs

- při porovnávání prezidentů bereme v úvahu nejen frekvenci tvarů, ale také jejich prominenci (DIN)
- pomocí DIN se zohledňuje působení na současného nativního mluvčího (čtenáře) – referenční korpus SYN2010

Řečnické strategie

Zaměření na gramatické a flektivní charakteristiky KWs

- při porovnávání prezidentů bereme v úvahu nejen frekvenci tvarů, ale také jejich prominenci (DIN)
- pomocí DIN se zohledňuje působení na současného nativního mluvčího (čtenáře) – referenční korpus SYN2010
- zastoupení gramatických kategorií v rámci klíčových slov – obvyklé gramatické charakteristiky prominentních slov

Řečnické strategie

Zaměření na gramatické a flektivní charakteristiky KWs

- při porovnávání prezidentů bereme v úvahu nejen frekvenci tvarů, ale také jejich prominenci (DIN)
- pomocí DIN se zohledňuje působení na současného nativního mluvčího (čtenáře) – referenční korpus SYN2010
- zastoupení gramatických kategorií v rámci klíčových slov – obvyklé gramatické charakteristiky prominentních slov
  - distribuce slovních druhů (subst., adj., verba, adv.)

Řečnické strategie

Zaměření na gramatické a flektivní charakteristiky KWs

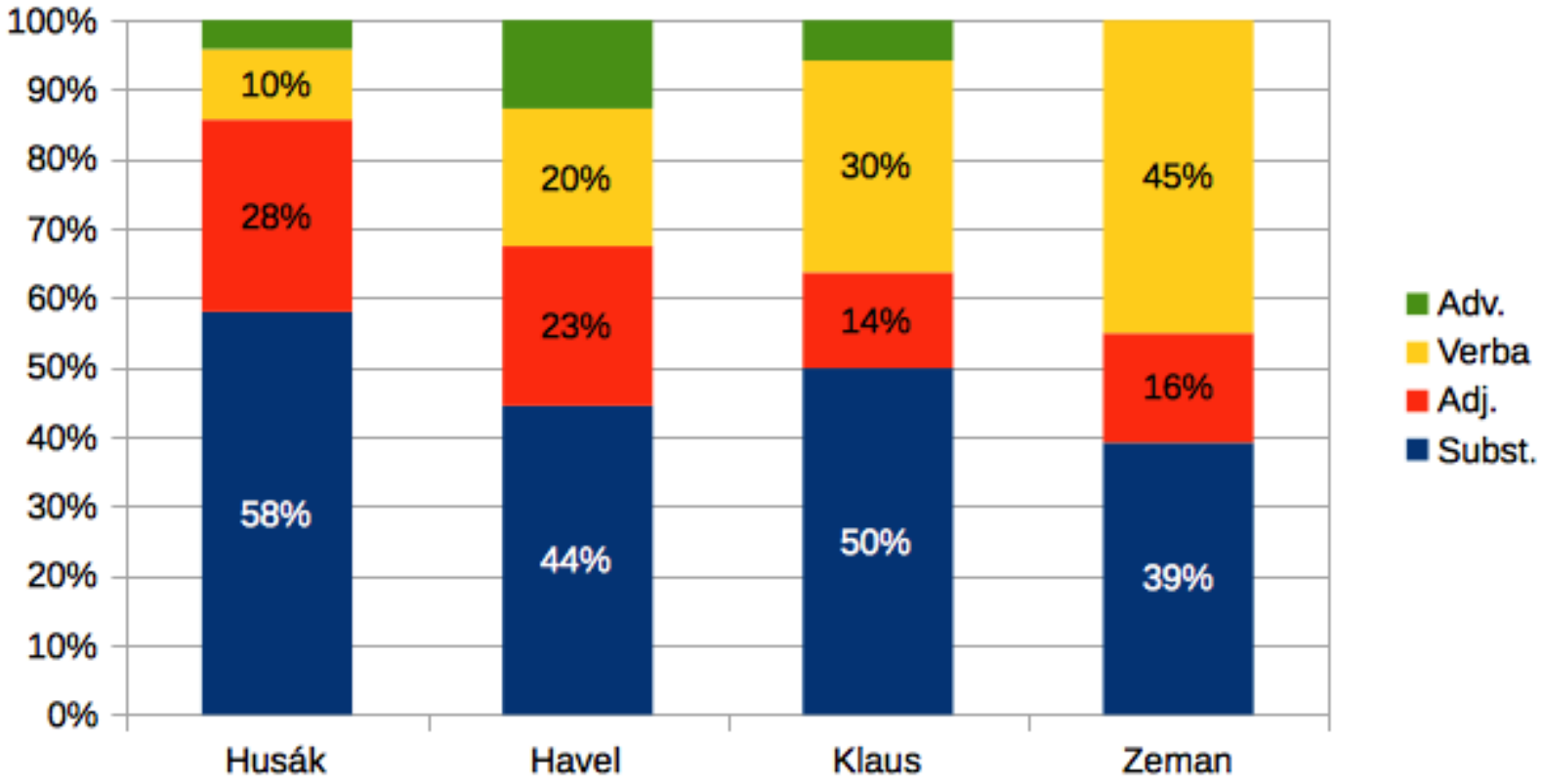
- při porovnávání prezidentů bereme v úvahu nejen frekvenci tvarů, ale také jejich prominenci (DIN)
- pomocí DIN se zohledňuje působení na současného nativního mluvčího (čtenáře) – referenční korpus SYN2010
- zastoupení gramatických kategorií v rámci klíčových slov – obvyklé gramatické charakteristiky prominentních slov
  - distribuce slovních druhů (subst., adj., verba, adv.)
  - distribuce slovesných osob

Řečnické strategie

Zaměření na gramatické a flektivní charakteristiky KWs

- při porovnávání prezidentů bereme v úvahu nejen frekvenci tvarů, ale také jejich prominenci (DIN)
- pomocí DIN se zohledňuje působení na současného nativního mluvčího (čtenáře) – referenční korpus SYN2010
- zastoupení gramatických kategorií v rámci klíčových slov – obvyklé gramatické charakteristiky prominentních slov
  - distribuce slovních druhů (subst., adj., verba, adv.)
  - distribuce slovesných osob
  - distribuce čísla u 1. osoby

Slovní druhy



Slovní druhy – interpretace

- slovesná vs. neslovesná (subst. + adj.) klíčová slova

Slovní druhy – interpretace

- slovesná vs. neslovesná (subst. + adj.) klíčová slova
  - Husák: prominence jmenných – statický, popisný projev

Slovní druhy – interpretace

- slovesná vs. neslovesná (subst. + adj.) klíčová slova
  - Husák: prominence jmenných – statický, popisný projev
  - Klaus (a Zeman) – důraz na akci (1. pl. imper. zejm. u VK)

Slovní druhy – interpretace

- slovesná vs. neslovesná (subst. + adj.) klíčová slova
  - Husák: prominence jmenných – statický, popisný projev
  - Klaus (a Zeman) – důraz na akci (1. pl. imper. zejm. u VK)
- podíl verbálních substantiv (např. *budování, plnění*)

Slovní druhy – interpretace

- slovesná vs. neslovesná (subst. + adj.) klíčová slova
  - Husák: prominence jmenných – statický, popisný projev
  - Klaus (a Zeman) – důraz na akci (1. pl. imper. zejm. u VK)
- podíl verbálních substantiv (např. *budování, plnění*)
  - Husák nejvíc (9 % ze subst.) – text explicitně neobsahuje čas, participanty ani modalitu

Slovní druhy – interpretace

- slovesná vs. neslovesná (subst. + adj.) klíčová slova
  - Husák: prominence jmenných – statický, popisný projev
  - Klaus (a Zeman) – důraz na akci (1. pl. imper. zejm. u VK)
- podíl verbálních substantiv (např. *budování, plnění*)
  - Husák nejvíc (9 % ze subst.) – text explicitně neobsahuje čas, participanty ani modalitu
- podíl adverbií – podrobnější kvalifikace děje

Slovní druhy – interpretace

- slovesná vs. neslovesná (subst. + adj.) klíčová slova
  - Husák: prominence jmenných – statický, popisný projev
  - Klaus (a Zeman) – důraz na akci (1. pl. imper. zejm. u VK)
- podíl verbálních substantiv (např. *budování, plnění*)
  - Husák nejvíc (9 % ze subst.) – text explicitně neobsahuje čas, participanty ani modalitu
- podíl adverbií – podrobnější kvalifikace děje
  - Havel (13 %) – modifikace potenciálně mylné představy posluchače, angažování publika

Slovní druhy – interpretace

- slovesná vs. neslovesná (subst. + adj.) klíčová slova
  - Husák: prominence jmenných – statický, popisný projev
  - Klaus (a Zeman) – důraz na akci (1. pl. imper. zejm. u VK)
- podíl verbálních substantiv (např. *budování, plnění*)
  - Husák nejvíc (9 % ze subst.) – text explicitně neobsahuje čas, participanty ani modalitu
- podíl adverbií – podrobnější kvalifikace děje
  - Havel (13 %) – modifikace potenciálně mylné představy posluchače, angažování publika
- podíl adjektiv – vytváření kontrastu subkategorizací

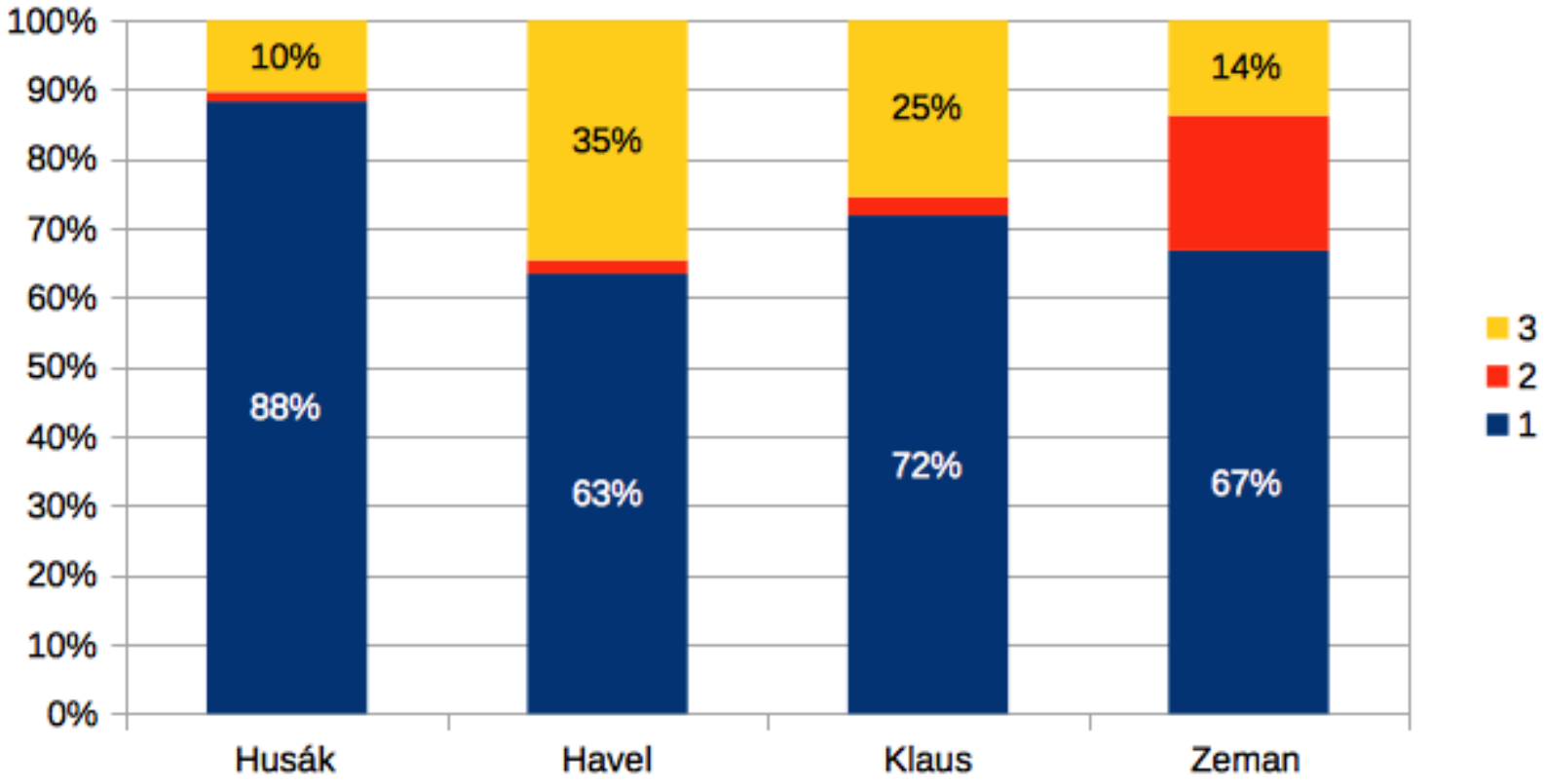
Slovní druhy – interpretace

- slovesná vs. neslovesná (subst. + adj.) klíčová slova
  - Husák: prominence jmenných – statický, popisný projev
  - Klaus (a Zeman) – důraz na akci (1. pl. imper. zejm. u VK)
- podíl verbálních substantiv (např. *budování, plnění*)
  - Husák nejvíc (9 % ze subst.) – text explicitně neobsahuje čas, participanty ani modalitu
- podíl adverbií – podrobnější kvalifikace děje
  - Havel (13 %) – modifikace potenciálně mylné představy posluchače, angažování publika
- podíl adjektiv – vytváření kontrastu subkategorizací
  - Husák (28 %) – socialistické státy vs. kapitalistický svět

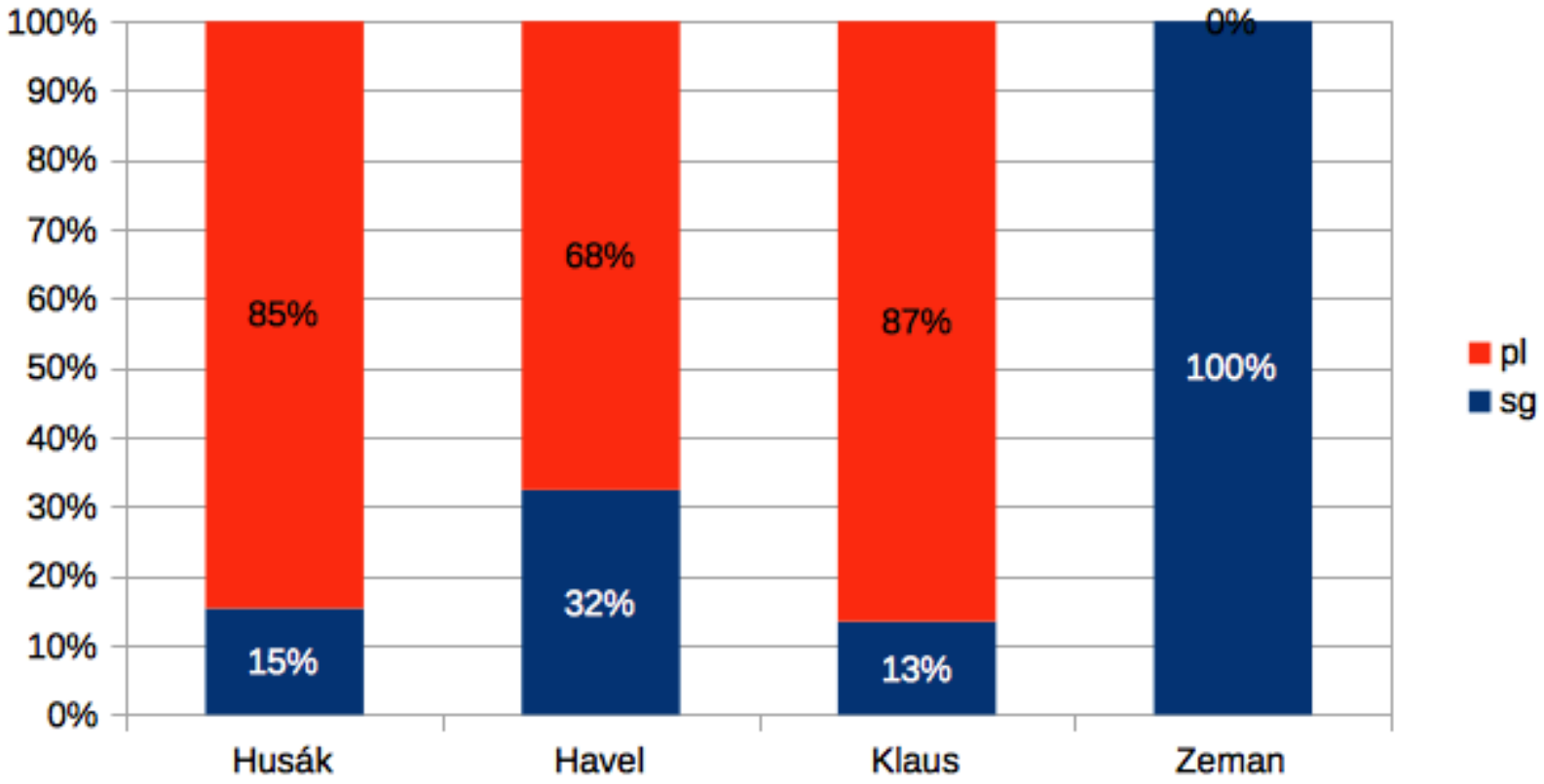
Slovní druhy – interpretace

- slovesná vs. neslovesná (subst. + adj.) klíčová slova
  - Husák: prominence jmenných – statický, popisný projev
  - Klaus (a Zeman) – důraz na akci (1. pl. imper. zejm. u VK)
- podíl verbálních substantiv (např. *budování, plnění*)
  - Husák nejvíc (9 % ze subst.) – text explicitně neobsahuje čas, participanty ani modalitu
- podíl adverbií – podrobnější kvalifikace děje
  - Havel (13 %) – modifikace potenciálně mylné představy posluchače, angažování publika
- podíl adjektiv – vytváření kontrastu subkategorizací
  - Husák (28 %) – socialistické státy vs. kapitalistický svět
  - Havel (23 %) – *ideální* kvalita/situace vs. *jiná* kvalita/situace

Slovesná osoba



Distribuce čísla – 1. os.



Slovesná osoba – interpretace

- 1. osoba je nejprominentnější (což lze v novoročním projevu očekávat – prezident jako zástupce celé společnosti)

Slovesná osoba – interpretace

- 1. osoba je nejprominentnější (což lze v novoročním projevu očekávat – prezident jako zástupce celé společnosti)
- největší podíl prominence má u Husáka, nejmenší u Havla

Slovesná osoba – interpretace

- 1. osoba je nejprominentnější (což lze v novoročním projevu očekávat – prezident jako zástupce celé společnosti)
- největší podíl prominence má u Husáka, nejmenší u Havla
- Husák klade největší důraz na 1. os. pl. (kolektivistické)

Slovesná osoba – interpretace

- 1. osoba je nejprominentnější (což lze v novoročním projevu očekávat – prezident jako zástupce celé společnosti)
- největší podíl prominence má u Husáka, nejmenší u Havla
- Husák klade největší důraz na 1. os. pl. (kolektivistické)
- Havel – pokud 1. os. používá – prominentněji mluví v sg. (explicitně prezentuje názory jako své vlastní)

Slovesná osoba – interpretace

- 1. osoba je nejprominentnější (což lze v novoročním projevu očekávat – prezident jako zástupce celé společnosti)
- největší podíl prominence má u Husáka, nejmenší u Havla
- Husák klade největší důraz na 1. os. pl. (kolektivistické)
- Havel – pokud 1. os. používá – prominentněji mluví v sg. (explicitně prezentuje názory jako své vlastní)
- z hlediska čísla v 1. os. můžeme sledovat podobný styl u Husáka a Klause

Charakteristika prezidentů

Pokus o charakteristiku jednotlivých řečnických strategií

- **Husák**: statický a deagentizovaný projev, vytváření kontrastu pomocí adjektiv, zahrnutí posluchače do sdělované reality (prominence 1. os. pl.), kolektivismus

Charakteristika prezidentů

Pokus o charakteristiku jednotlivých řečnických strategií

- **Husák**: statický a deagentizovaný projev, vytváření kontrastu pomocí adjektiv, zahrnutí posluchače do sdělované reality (prominence 1. os. pl.), kolektivismus
- **Havel**: středně dynamický typ, zájem o okolní svět (3. os.), důraz na kvalifikátory (adj. i adv.), individualismus ve vlastním vystupování (1. os. sg.) i v referování o okolním světě (3. os. sg.)

Charakteristika prezidentů

Pokus o charakteristiku jednotlivých řečnických strategií

- **Husák**: statický a deagentizovaný projev, vytváření kontrastu pomocí adjektiv, zahrnutí posluchače do sdělované reality (prominence 1. os. pl.), kolektivismus
- **Havel**: středně dynamický typ, zájem o okolní svět (3. os.), důraz na kvalifikátory (adj. i adv.), individualismus ve vlastním vystupování (1. os. sg.) i v referování o okolním světě (3. os. sg.)
- **Klaus**: nejdynamičtější projev, stejné využití 1. os. pl. jako u Husáka (kolektivismus, vystupování jako zástupce celé společnosti), nejmenší potřeba kvalifikovat

Charakteristika prezidentů

Pokus o charakteristiku jednotlivých řečnických strategií

- **Husák**: statický a deagentizovaný projev, vytváření kontrastu pomocí adjektiv, zahrnutí posluchače do sdělované reality (prominence 1. os. pl.), kolektivismus
- **Havel**: středně dynamický typ, zájem o okolní svět (3. os.), důraz na kvalifikátory (adj. i adv.), individualismus ve vlastním vystupování (1. os. sg.) i v referování o okolním světě (3. os. sg.)
- **Klaus**: nejdynamičtější projev, stejné využití 1. os. pl. jako u Husáka (kolektivismus, vystupování jako zástupce celé společnosti), nejmenší potřeba kvalifikovat
- **Zeman** (nedostatečná data): dynamický projev, zaměřený na posluachače (2. os.) a na sebe (1. os. sg.)

Děkuju za pozornost!