



# Usage-based Approach to Discourse through Keyword Analysis

Masako Fidler (masako\_fidler@brown.edu) and Václav Cvrček (vaclav.cvrcek@ff.cuni.cz)

Brown University, USA & Charles University in Prague, Czech Republic

ICLC at the U. of Alberta, Edmonton, Canada (June 23-28, 2013)

## Theoretical objective

Expansion of the usage-based approach (Hopper 1987, Bybee 2010) to the level of overarching themes in discourse.

## The main idea

**Contrast in word frequencies** (target text vs. reference corpus) → **Contrast in patterns of language use** between the two →

**Special properties of the text** (“keyness”) (Scott and Tribble 1996, Baker and Ellece 2011) →

**We may be able to unpack changing properties of the society in political discourse over time.**

## Materials

**Target texts:** fifteen New Year's Addresses in Czech from 1975 to 1989 by President of the totalitarian Czechoslovakia ≈1, 000 words each

**Reference corpus:** Periodicals from 1952-1969

**Motivations for using this material:**

- Control (same author, same genre)
- Texts from a totalitarian period do not openly discuss societal changes

**Method: “Needle in a Haystack” Corpus-based detection of thematically prominent word forms**

**“KWords” software** (<http://kwords.korpus.cz/>) **processes a text and provides a list of prominent words (keywords) - Keywords are NOT query words; they are outputs!**

**Main features of the software:**

- thematically prominent words culled and ranked by DICE’s co-efficient (KWs)
- thematic links and their visualization (pic. on the right) (KW links)
- collocation for each KW intended for detecting subtle shifts in topics in the same genre over time (hence the name Needle in a Haystack project)
- comparison of KW rankings over time

**Data for discussion**

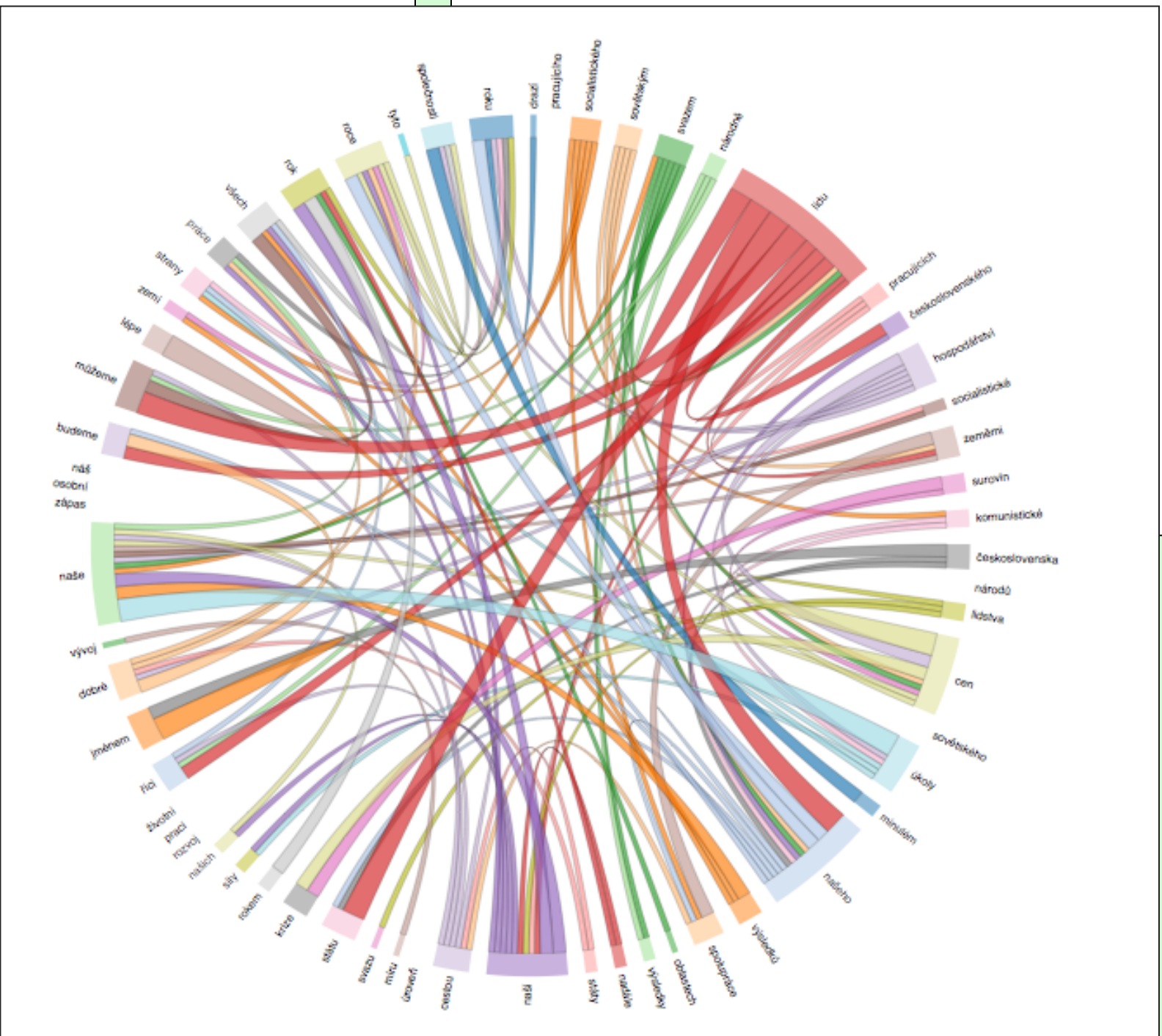
**Recurring/frequent thematic word forms**

(keyness and KW links over time)

**Near-synonyms** (keyness over time)

(terms of address, names for the same political entity)

**Isolated KWs:** KWs that occur once or sporadically



## Correspondences are found between the data and important historical events:

**Graph 1: Recurrent/frequent KWs: types of 1<sup>st</sup> pers. (personal and possessive) pronominal forms that were picked as KWs (1975-1989)** (“us, our” (nás, náš, naše, našeho, našem, naši, naší, našich, našim) (larger numbers indicate keyness spreads out over more diverse syntactic functions) 1980 (demonstrations in the neighboring Poland), 1982 to 1985 (instability in the Soviet leadership), 1989 (increasing unrest in Czechoslovakia) 1981, 1982 (the government crackdown on the Polish demonstrations); 1986-87 (provisional stability in the Soviet leadership)

**Graph 2: Different use of apparently near-synonymous KWs**

- “friends!” (přátelé!) ≈ “comrades!” (soudruzi!) (1978, immediately after Charter 77) whereas “friends!” ↔ “comrades!” (elsewhere)
- “comrades!” (and not “friends!”) increase in keyness (the Eastern Bloc regains control)
- “co-citizens!” (spoluobčané!) sometimes lines up with “comrades” (76, 79, 80, 81, 83, 84, 85, 86, and 88), but its keyness is a mirror image of “comrades!” (soudruzi) in 1977-1978 and 1982 & 1987 (politically more precarious years)

**Graph 3: Recurrent KWs – the number of politically-related KW links to frequent KW “of people” (lidu)**

- appearance of plurality: link to “of the (National) front” in the late 1970s (signing of the Helsinki Accords) and 1985(beginning of glasnost)
- shifts in the main concern – link to state “státu” and party “strany” in the late 1970s vs. homeland (and society) increasing in the strength of links towards the 1980s (Charter 77 human rights movement, demonstrations in Poland)

**Isolated KW** appear in specific periods marked by domestic and international events.

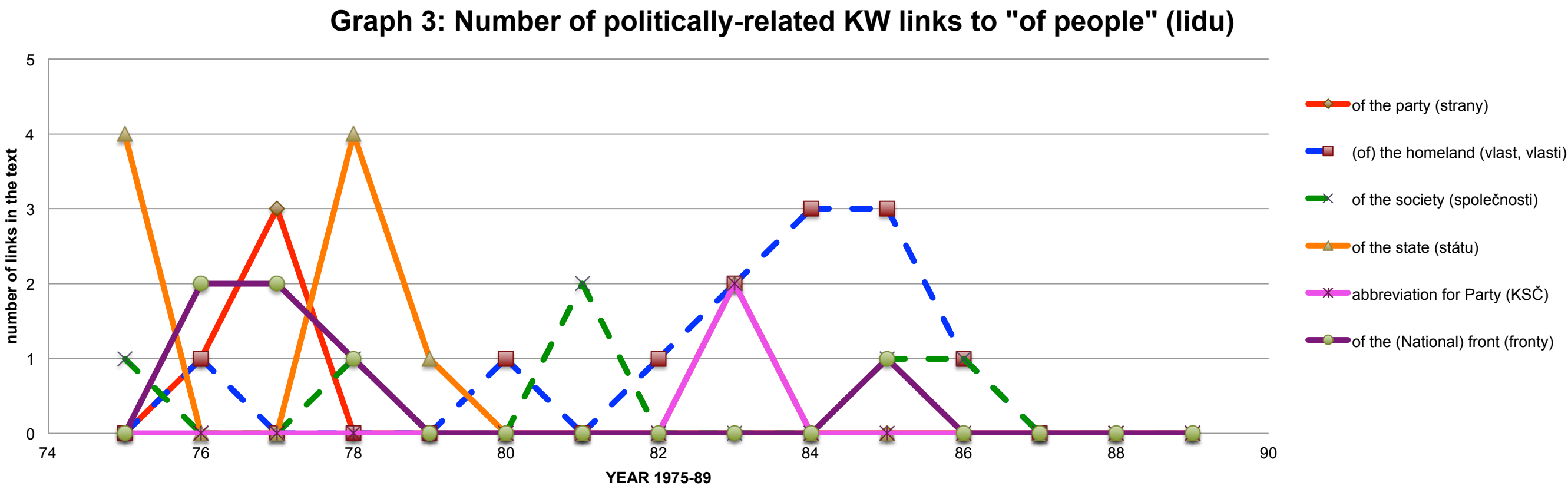
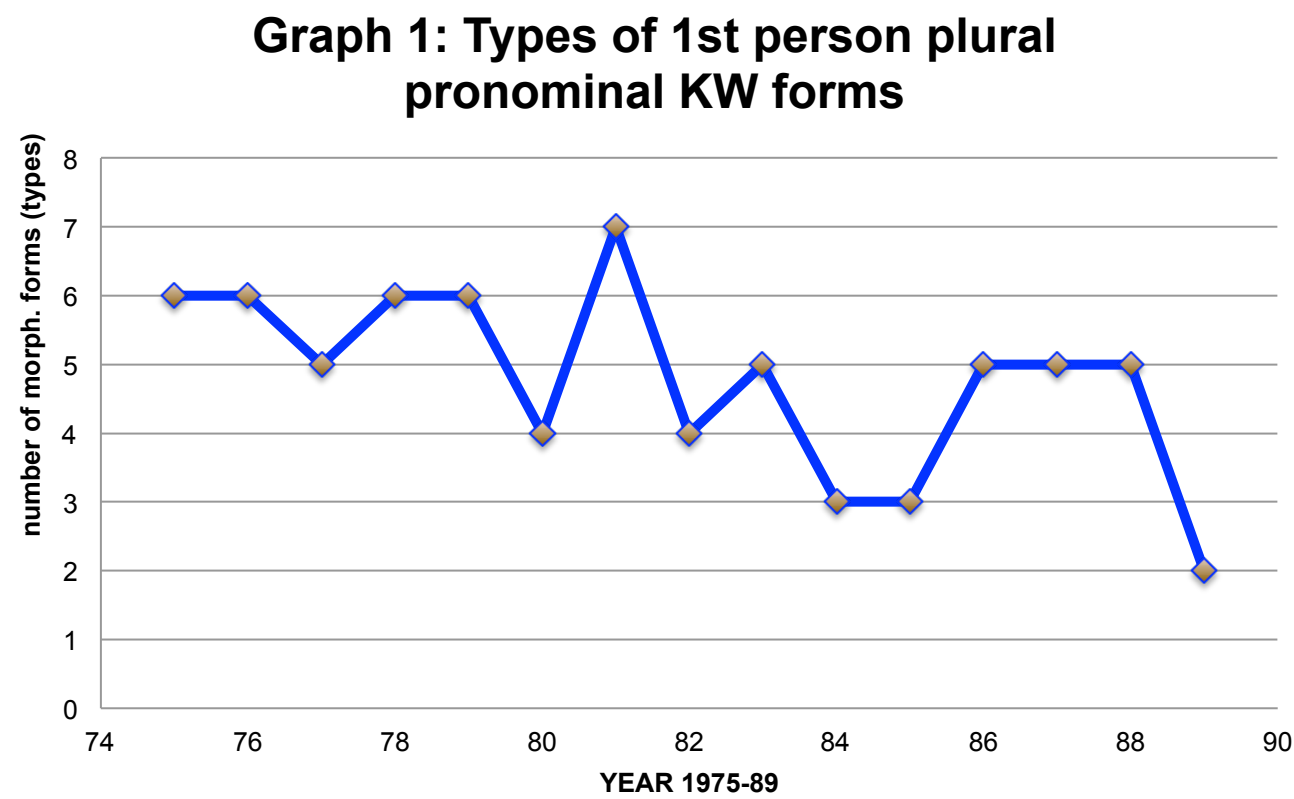
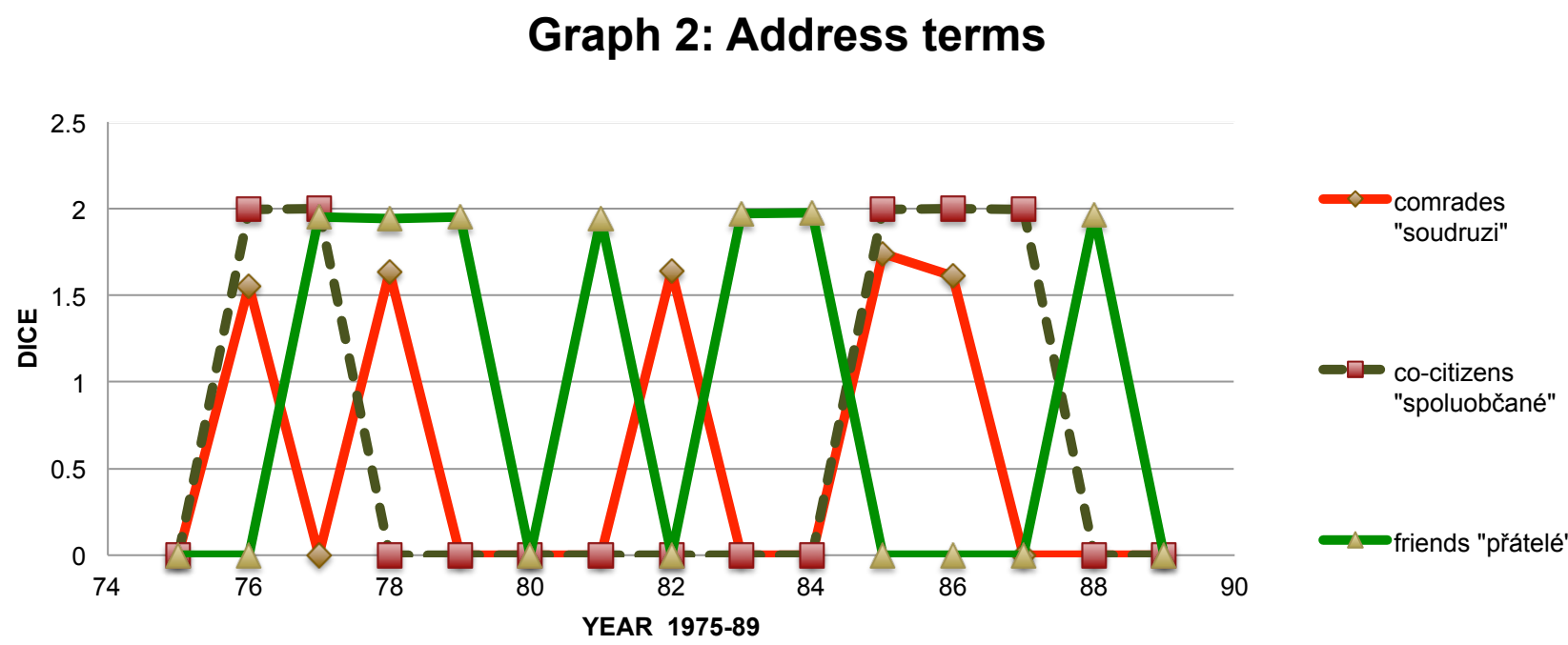
1975: (with) the USSR, natural resources, prices, crisis, humankind (world-wide economic crisis, oil shock)

1976-79: social structure, “layers” (pseudo-equivalent of social classes) (Czechoslovak government to sign the Helsinki Accords, Charter 77 human rights movement)  
1980-81: of peace (adjective forms), tension, (international) relations (Unrest in Poland, US sanctions)  
1982-86: peace (noun), disarmament, politics (noun and adjective forms), position (of Czechoslovakia), powers (the government crackdown on the Polish demonstrations, alternating leadership in the USSR till 1985, arms talk between Gorbachev and Reagan)

## References

Bybee, J. 2010. Language, usage and cognition.  
Baker, P. and Ellece, S. 2011. Key terms in discourse analysis. The Czech National Corpus ([www.korpus.cz](http://www.korpus.cz)).  
Hopper, Paul. Emergent grammar. Berkeley Linguistics Society 13: 139-157. Berkeley Linguistics Society, vol. 13 (1987), 139-157  
Scott, M. and C. Tribble. 2006. Textual patterns: Keyword and corpus analysis in language education.

## Results (highlights)



## Discussion and conclusions

This corpus-based analysis of prominent word forms indicates subtle reactions of the Czechoslovak leadership to the major events in history.

Statistical contrast in language use (target text vs. ref. corpus) reflects topic shifts in texts these topic shifts in turn signal subtle societal changes over time.

The following aspects of corpus-based keyword analysis help shed light on topic and societal changes in text over time.

- Ranking of KWs (degrees of keyness)**
- Diversity of syntactic functions to which keyness is attributed**
- Changing set of KWs and KW links**

**For more discussion see:**

Fidler-Cvrček, [https://documents.ku.edu/xythoswfs/webview/fileManager?stk=&entryName=%2Fusers2%2Fmlg%2FSL5\\_7\\_handouts&msgStatus](https://documents.ku.edu/xythoswfs/webview/fileManager?stk=&entryName=%2Fusers2%2Fmlg%2FSL5_7_handouts&msgStatus) (look under Fidler-Cvrček)  
Cvrček-Fidler, [kwords.korpus.cz/texty/cads2012.pdf](http://kwords.korpus.cz/texty/cads2012.pdf)  
Fidler-Cvrček, <http://www.aatseel.org/program/2013-aatseel-conference-program>  
Cvrček-Fidler, <http://ucrel.lancs.ac.uk/cl2013/programme.php> (abstract to be published)

**Acknowledgments:**

Pavel Vondříčka, The Institute of the Czech National Corpus at Charles University in Prague, Humanities Research Funds from Brown University, Travel funds from the Office of International Affairs, Brown University