Masako Fidler* and Václav Cvrček

Keymorph analysis, or how morphosyntax informs discourse

Abstract: This paper introduces keymorph analysis (KMA), a new extension of the discourse-probing technique of keyword analysis (KWA). While KWA focuses on lexicon and provides a key predominantly to textual topics and their semantic associations, KMA focuses on morphosyntactic features and captures more general characteristics of texts as wholes. Speeches by Czech(oslovak) presidents and corpus of contemporary written Czech SYN2015 are used to illustrate readers' perception of situations and their participants in these texts. The paper also shows that different levels of morphosyntax facilitate interpretation of discourse: ideological differences can be observed with higher-level morphosyntactic features (parts of speech), while differences in speaker style are observable with lower-level morphosyntactic features (case, number, person, and their combinations).

Keywords: keyword analysis, morphosyntax, cognitive linguistics, discourse, Czech

1 Introduction

Existing literature demonstrates grammar–discourse interaction. Hopper and Thompson (1980) find correlation between the discourse properties of foregrounding–backgrounding and grammatical components with high–low "Transitivity". Quantitative studies by Biber (1993, 2006) show how lexical and morphosyntactic data facilitate the identification of linguistic registers in English. Kresin (1998) shows how choice of deixis reflects thematic hierarchization in discourse. Studies in Slavic languages describe how verbal aspect facilitates the organization of discourse (Chvany 1990; Fielder 1990; Sonnenhauser 2008; and Altshuler 2010; on Russian; Desclés and Guentschéva 1990; on Bulgarian).

*Corresponding author: Masako Fidler, Department of Slavic Studies, Brown University, Providence, RI, USA, E-mail: masako_fidler@brown.edu
http://orcid.org/0000-0002-3474-8581
Václav Cvrček, Faculty of Arts, Charles University, Institute of the Czech National Corpus, Prague, Czech Republic, E-mail: vaclav.cvrcek@ff.cuni.cz
http://orcid.org/0000–0003–3977–2393

While these extensive studies focus on the co–occurrence of grammatical (and lexical) components and discourse functions and types, this paper uses morphosyntactic features that diverge from the general language usage pattern and by doing so attempts to explore aspects of discourse that are expected to be noteworthy to the contemporary interpreter. Our method extracts those morphosyntactic features that help shape the average reader's interpretation of a text.

By morphosyntactic features, we mean identifiable[1] inflectional markers (person, number, case, verbal negation, verb finiteness) and parts of speech (POS). Analysis of these features is built on the cognitive linguistic view that linguistics units and categories are meaning–bearing in all contexts (Divjak and Janda 2008: 139). These morphosyntactic features are expected to reflect how speakers accentuate or de–accentuate parts of the world in discourse, thereby making the discourse accessible for the addressee to confirm, contest, or question the specific components in the text. The process of (de–)accentuating features out of a range of available choices (in both verbal text and images) belongs to the "representational strategies" of Critical Discourse Analysis (Fowler 1991; Van Dijk 1993; Fairclough 2003; Wodak 2009).

Our discussion is based on what we term keymorph analysis (henceforth KMA). KMA uses the principles of keyword analysis (KWA), a corpus linguistic method. KWA focuses on lexicon (words, lemmas and/or phrases), revealing prominent topics and semantic associations, which affect the discourse images of individuals, institutions, and situations (Culpeper 2002; Scott and Tribble 2006; Baker 2009). The extracted data (keywords) are connected with what the text is about and its stylistic properties (Scott 2010: 43). Keywords can also show the average reader's perception of what is striking in a text when a target text is contrasted with a robust and well–balanced reference corpus that reflects language use of the given time (Fidler and Cvrček 2015). Application of the principles of KWA to grammatical information, however, has been limited to parts of speech, as in Culpeper (2009)'s Key POS analysis.

In contrast, KMA probes the relationship between discourse and morphosyntactic features used in text[2]: not only parts of speech, but also more detailed

---

1 We therefore also include inflectional forms that do not possess a morphologically distinct form (e. g. voc pl of Czech nouns), which can be automatically disambiguated.

2 As the focus of this study is the probable *reception* of these texts rather than the way they were produced, issues concerning the authorship of these texts (e. g. the presence of ghost-writers and censorship) is outside the scope of the discussion.

morphosyntactic features. The highly inflective nature of the data in Czech facilitates our understanding of how morphosyntax can inform discourse and complements the existing qualitative and quantitative text analysis.[3]

The following section presents the data (Section 2). Sections 3 and 4 discuss the methodology, including the precautionary measures taken against potential pitfalls as well as the grammatical features excluded from the analysis. These sections are followed by the interpretation of results (Sections 5–7) and the conclusions (Section 8).

2 Data

The data are drawn from SPEECHES[4] (Cvrček et al. 2015), a 249,000 word corpus containing Czechoslovak and Czech presidential addresses,[5] delivered on anniversaries and holidays from 1918 to 2014. The speeches are from four distinct political periods: pluralistic democratic presidencies, presidency without state sovereignty (marked with dark shading in all the tables), totalitarian socialist presidencies (marked with light shading in all the tables), and post–socialist pluralistic democratic presidencies (Figure 1).[6] The texts are prototypical examples of prepared and monologic formal discourse in a written–to–be–spoken mode.

Most of the speeches were delivered on New Year's Day (more than 51 % of tokens) and on Republic Day (October 28) (34 %). The others are Christmas messages (13 %) and speeches on Protectorate Day (1 %, only during Hácha's presidency).[7] All texts[8] were lemmatized, morphologically tagged, and enriched by metadata.

---

3 Těšitelová (1995) discusses frequencies of parts of speech, morphosyntactic forms, and lexicon of President Beneš, but the data are specific to one author without the use of the RefC.
4 This corpus, which resulted from the cooperation between the Czech National Corpus and the University of Oslo, is available at www.korpus.cz via the KonText search engine.
5 Or addresses by politicians who spoke on behalf of a president if the post was vacant or during the president's illness (e. g. Husák speaking on behalf of president Svoboda in 1975).
6 For the sake of space, we use the terms "socialist" and "democratic" in the remainder of the paper.
7 Speeches delivered by speakers other than those listed in Figure 1 were excluded from the research.
8 The following notations are used in citations: n (New Year's Day), r (Republic Day), v (Christmas message), and p (Protectorate Day).

President	Abbreviation	In office	Tokens	Speeches	Notes on the speaker and/or texts
Tomáš Garrigue Masaryk	TGM	1918–1935	22,385	17	The first president ("the founding father") of Czechoslovakia.
Edvard Beneš	EB	1935–1938; 1945–1948	45,773	21	His speeches from exile (1939-1945) were also included.
Emil Hácha	EH	1938–1945	5,767	13	President of the Protectorate of Bohemia and Moravia under the Nazi Occupation.
Klement Gottwald	KG	1948–1953	11,390	6	The first communist president.
Antonín Zápotocký	AZ	1953–1957	10,819	4	The presidency starts with Stalin's death. Khrushchev's secret speech at the 20th Congress of the Soviet Communist Party denounced Stalin in 1956, one year before the president died in office.
Antonín Novotný	AN	1957–1968	30,161	11	President in the post-Stalin era in Czechoslovakia.
Ludvík Svoboda	LS	1968–1975	11,421	7	The presidency starts at the peak of the Prague Spring reform movement and continues after the Soviet invasion in 1968. The New Year's Address in 1975 was delivered by Husák.
Gustav Husák	GH	1975–1989	27,374	17	The last communist president during the post-invasion ("Normalization") period.
Václav Havel	VH	1989–2003	47,853	25	The first democratic president of Czechoslovakia and the Czech Republic after the "Velvet" revolution.
Václav Klaus	VK	2003–2013	26,650	21	A center-right conservative president, known for his euro-skepticism and anti-global-warming view.
Miloš Zeman	MZ	2013–present	5,086	4	The corpus contains only texts from the period 2013–2014.
TOTAL			244,679	146	

Figure 1: Corpus "SPEECHES".

3 Methodology

Results were extracted by comparing the speeches of a single president (the target text or corpus, henceforth Ttxt) to the entire SPEECHES (the reference corpus, henceforth RefC). In addition, we used SPEECHES as the Ttxt and SYN2015 as the RefC to identify the genre–specific features of presidential speeches. SYN2015 is a 100 million–word corpus that represents the majority of language variation (i. e., written, public, and printed discourse) (Křen et al. 2016); it is possible to regard this corpus as reflecting a large spectrum of text types, i. e., the language patterns to which the average contemporary reader in Czech is likely to be routinely exposed.

When comparing frequency characteristics, two types of statistical information are crucial in assessing the prominence of an item in a text: (1) statistical significance, which confirms that the amount of data available for comparison is sufficient, and (2) the effect size of the difference, which informs about the *prominence* of the linguistic item. The former is usually examined by the chi2 test or the log–likelihood test (which this study employs), while the latter can be estimated with Difference Index (DIN), an effect size estimator, which compares the difference between relative frequencies of an item in the Ttxt and the RefC with the mean of those relative frequencies. The value (multiplied by 100) ranges from −100 (for items nonexistent in the Ttxt), to 100 (for items occurring in the Ttxt only). The zero value of DIN appears when an item is equally evidenced in the Ttxt and the RefC (cf. Fidler and Cvrček 2015).

This study identifies prominent morphosyntactic features or keymorphs, using the same principle used in KWA (Scott and Tribble 2006; Culpeper and Demmen 2015): the relative frequency of a morphosyntactic feature in the target corpus is compared with the relative frequency of the same feature in the RefC. DIN values alone, however, are not sufficient in KMA.[9] KMA requires the notion of "noteworthiness" because keymorphs differ from keywords in two aspects: in their size of repertoire and their inherent hierarchical structure. They must be discussed before we analyze data.

---

9 The terms "keymorph" and "keymorph analysis" parallel the terms "keyword" and "keyword analysis". A "key*word*" in KWA may refer to a lemma and/or a word form: e. g. the abstract representation of all the inflected forms of the noun *hrad* [castle] and/or the surface realizations (inflected forms) of that lemma *hrad, hradu, hrade, hradem, hradě, hrady, hradů, hradům, hradech*). Similarly, a key*morph* can refer to a morpheme and/or morphs; e. g. the morpheme {1sg non–past} and/or the surface representations of the morpheme, such as suffixes {ám}, {ím}, {u}. The current study examines the former.

3.1 Importance of repertoire size

Unlike in KWA, the size of the repertoire in KMA affects the overall frequency dispersion of the phenomena under scrutiny. This influences the traditional methods of obtaining statistical significance and effect size, while also bringing new interpretive challenges. While words or lemmas are items of low to moderate frequency in KWA, morphosyntactic features can be items of low to *very high frequency* in KMA. In fact, some of these features are so frequent that they can be described as core constituents of a text.[10] A typical grammatical feature thus usually yields a statistically significant difference between frequencies in the Ttxt and the RefC. This does not inform us about the *prominence* of a feature; it simply means that we have enough data to infer that the difference (however small) is unlikely to be caused by pure chance.[11] Therefore, especially when dealing with high frequency phenomena such as POS, it becomes crucial that we should not conflate statistical significance and the effect size of the difference.[12]

The prominence level of a morphosyntactic feature can also vary in accord with its overall frequency. High–frequency keymorphs are already so predominant in language that they may depart only slightly from a DIN value of zero (denoting equal representation of an item in the Ttxt and RefC) – yet a relatively low but still positive DIN might be quite informative of the nature of discourse.

The dispersion of frequencies of morphosyntactic features is unproblematic in inter–*speaker* comparison: when comparing how different speakers use the same morphosyntactic feature.[13] When we compare DIN values of a single morphosyntactic feature used by different speakers/presidents, we compare them against the same background: the frequency of this feature in the RefC. Inter–speaker comparison is therefore possible, even though average DIN values for various morphosyntactic features may vary (e. g. verbs might score higher than nouns or pronouns).

We must, however, use caution in inter–*feature* comparison, i. e. when attempting to characterize the overall nature of a single speaker's discourse using more than one keymorph, since not every morphosyntactic feature plays

---

10 For example, nouns (as a grammatical category) represent 1/4 of the whole corpus of SPEECHES, whereas some uncommon morphosyntactic features, such as 2nd pl ind pres verb forms, are represented only by 170 occurrences (0.07 % of the whole corpus).

11 Unless otherwise stated, all differences in frequencies of features reported in this paper are statistically significant at the level $p < 0.05$ (measured by log–likelihood test).

12 Cf. Fidler and Václav (2015) for detailed discussion on the problem in conflating statistical significance and effect size.

13 This mode of analysis involves horizontal reading of a table, comparing DIN values of presidents (cf. Table 4) for one grammatical category (a single row).

an equally important role in discourse.[14] It is at least questionable for a feature with a high DIN value to outperform a feature of a moderate (but still positive) DIN value, when the former occurs with a very low frequency and the latter with a very high frequency. Consider a feature such as the 2pl indicative present verb forms, which may be used only 5 times per text in this genre. If a speaker uses this feature 10 times (i. e. twice as much as expected), his DIN value for this feature could be approximately 33. In contrast, if the same speaker uses a very frequent feature such as nouns in the same text at the frequency of only 5 % above the expected value, his/her DIN for nouns would be only 2, since nouns are much more frequent. In such a situation, nouns, despite their lower DIN as a part of speech, are expected to contribute to discourse more than the 2pl indicative present verb forms because of their considerably higher overall frequency.

For a methodologically sound inter–feature comparison and for ease of constructing a description of presidential discourse strategies, we divided all the potentially useful features into five classes according to their overall frequencies in SPEECHES[15]: core constituents of the text, frequent features, common features, sporadic features, and exceptional features. Table 1 is the list of those features included in this paper.[16]

Table 1: Morphosyntactic features used in this study and their classification.

Feature	SUM	Ipm (instances per million)	Log$_2$ (ipm)	Class	Label
All tokens	244,679	1000,000	19.9	5	Within the whole corpus (SPEECHES)
Nouns and adjectives	90,847	371,290.5	18.5	5	core constituent of the text
Nouns (without adjectives)	59,341	242,525.9	17.9	5	core constituent of the text
Verb finite forms	22,382	91,475.0	16.5	5	core constituent of the text
Nouns, genitive case	19,846	81,110.4	16.3	5	core constituent of the text
Nouns, accusative case	12,317	50,339.4	15.6	4	frequent feature
Nouns, nominative case	10,441	42,672.2	15.4	4	frequent feature

(*continued*)

---

14 This mode of analysis involves vertical reading of a table (cf. Table 4), identifying two or more prominent grammatical features observed in one speaker's texts, and describing his/her discourse strategy.

15 The scale is derived from the base 2 logarithm of instances per million (ipm) divided into 5 ranges.

16 The features of the last category (exceptional features) were not used for the current analysis. The complete list of Classes is available on request.

Table 1: (*continued*)

Feature	SUM	Ipm (instances per million)	Log₂ (ipm)	Class	Label
Possessive pronouns	6,758	27,619.9	14.8	4	frequent feature
Nouns, instrumental case	5,731	23,422.5	14.5	4	frequent feature
Possessive pronoun–*náš* 'our'	3,786	15,473.3	13.9	4	frequent feature
Personal pronouns	3,704	15,138.2	13.9	4	frequent feature
Nouns, dative case	3,198	13,070.2	13.7	4	frequent feature
Verbal negation	2,671	10,916.3	13.4	4	frequent feature
Verbs indicative present 1pl	2,214	9,048.6	13.1	4	frequent feature
Personal pronoun – *my* 'we'	1,599	6,535.1	12.7	4	frequent feature
Verbs indicative present 1sg	1,349	5,513.3	12.4	4	frequent feature
Personal pronoun – *já* 'I'	297	1,213.8	10.2	3	common feature
Possessive pronoun – *můj* 'my'	137	559.9	9.1	3	common feature

While DIN values can in principle serve as a basis for comparing speakers/ presidents and for identifying their prominent features, these classes are necessary to estimate the importance of a feature *in relation to* the other features used by the same speaker/president. For example, a high DIN value of a sporadic feature is not expected to have the same impact on discourse as the same DIN value of one of the core constituent features.

3.2 Overt hierarchical structure of morphosyntactic features

The second specificity of KMA concerns the hierarchical relationship overtly present in grammar. By "hierarchical" we mean that morphosyntactic features are interconnected: e. g. nouns mark case in Czech; consequently, the prominence of a specific case (a lower–level morphosyntactic feature) may be a mere consequence of the prominence of nouns as a whole (a higher–level morphosyntactic feature). Thus, in texts by a speaker who uses nouns more than average, all nominal forms with different case endings may be prominent; similarly, a speaker who overuses verb forms may overuse all lower–level verb forms such as the indicative forms, conditional forms, and imperative forms.

So as not to lose sight of any possible interdependency among morphosyntactic features on different levels, we have enriched the description of each feature under examination with additional information: the correlation between the higher–level parts of speech (POS) and the lower–level morphosyntactic feature.[17] A stronger correlation between a morphosyntactic feature and a POS suggests that the prominence of the former *might* be influenced or caused by the prominence of the latter, while a weaker correlation indicates a more autonomous status of the morphosyntactic feature independent of the POS (at least in this genre).

Table 2 summarizes the correlation coefficients of verbal, nominal and pronominal features that are relevant to the analysis (Sections 5–7).

Table 2: Correlation between POS and the morphosyntactic features used in this study.

Higher–level morphosyntactic feature: Part of speech (POS)	Lower–level morphosyntactic feature	rho	Interpretation
Verbs	finite verb forms	1.00	100 % correlation
	verbal negation	0.86	very strong
	1sg indicative present forms	0.86	very strong
	1pl indicative present forms	0.89	very strong
Nouns	nominative case	0.90	very strong
	genitive case	0.97	very strong
	dative case	0.98	very strong
	accusative case	0.99	very strong
	instrumental case	0.97	very strong
Pronouns	personal pronouns	0.94	very strong
	possessive pronouns	0.96	very strong
	1sg personal pronouns	0.50	moderate
	1pl personal pronouns	0.95	very strong
	1sg possessive pronouns	0.32	weak
	1pl possessive pronouns	0.95	very strong

The correlation coefficient here indicates the overall tendency for many of the lower–level features to be influenced by the frequency of the higher–level POS. Clearly, interpretation of many of these lower–level morphosyntactic features requires an additional procedure: to determine whether they are "noteworthy". Noteworthiness is an umbrella term for conditions relevant to the proper

---

17 The data–points were frequencies of grammatical categories for each speaker in corpus SPEECHES; as the data are partly categorical, Spearman's rank correlation coefficient (rho) was used.

interpretation of prominent morphosyntactic features, which will be discussed in the following section.

3.3 Noteworthiness

Noteworthy results are those results that a researcher can *safely* consider as relevant in his/her assessment and interpretation of the characteristics of discourse. For those morphosyntactic features that are not strongly correlated with others (e. g. POS), two criteria below are sufficient to identify noteworthy results:
1. The frequency difference of the feature in the Ttxt and RefC is statistically significant.
2. The DIN value (the effect size of the difference) is positive.

In Table 4 (Section 5), e. g., the results in boldface are noteworthy; they are both statistically significant and show positive DIN values.

When a morphosyntactic feature is strongly correlated with a higher–level feature (as seen in Table 2), an additional criterion is necessary to ensure that we examine results that are noteworthy:
3. The DIN value of the lower–level morphosyntactic feature must be equal to or higher than the DIN value of the relevant higher–level feature.

This criterion is necessary to address a potential *halo effect* in which a lower–level feature has a high DIN value not due to its own prominence, but as *a possible consequence* of the prominence of a related higher–level feature (cf. 3.2). To illustrate, the DIN value for the genitive case in speeches (Table 6 in Section 6.2) by GH (21.56) is deemed noteworthy because the difference is statistically significant, positive, and the DIN value is higher than the DIN of nouns (POS) (12.01). Even when the DIN value is positive and is larger than the DIN value for the higher related morphosyntactic feature, the result is not noteworthy unless it is statistically significant; e. g., the DIN value for the instrumental case in speeches by MZ (5.88 in Table 6) is not noteworthy although it is positive and larger than the DIN value for the nouns (POS) (–2.22) because it does not meet the condition of statistical significance.[18]

---

18 Positive DIN values that are not statistically significant are presented in normal (i. e., non–bold) style in tables throughout this paper. The table showing statistical significance of morphosyntactic features in presidential texts can be obtained from the authors upon request.

4 Genre–specific features that were excluded from analysis

Genre–specific morphosyntactic features typical of presidential speeches do not illuminate individual speaker styles and they must be excluded from the analysis. These features are listed in Table 3. The vocative pl nouns are highly frequent as presidents usually address the nation rather than an individual. The vocative sg nouns and the 2sg verb finite forms[19] are rare as the genre does not allow forms typical of speech based on informal relationships. The table below confirms this expectation. As these features reflect nothing new but the typical manifestation of genre (rather than speaker differences), they are not considered for further analysis.[20]

Table 3: Genre–specific morphosyntactic features excluded from analysis.

Morphosyntactic forms	DIN within SPEECHES
Nouns in the vocative pl[a]	93.67
Nouns in the vocative sg	−38.56
2sg verb forms	−89.09

RefC: SYN2015
[a]The vocative pl. forms in Czech are homonymous with the nominative pl., but they are signaled by punctuation and can be disambiguated with the help of syntactic constraints (the presence of two nominative forms; the frequent co–occurrence with the predicate in 2pl forms).

Sections 5–7 below will discuss other morphosyntactic features that reflect characteristics of speakers and their ideological orientations.

5 KMA: Parts of speech (POS)

POS are morphosyntactic features on the highest level. Some of them point to differences in discourse properties among the presidents. Below are the DIN

---

19 By finite we mean verb forms that explicitly mark the grammatical person. This category includes the imperative mood forms in Czech.

20 The presence or absence of the 2pl forms is not as typical of the genre as the forms presented in this table. A preliminary analysis of the use of these forms suggests that the presidents in transitional periods (TGM, EB, AN, LS, and VH) show more frequent use. Prominence of this feature may signal the degree to which the speakers engage the audience (foundation of the first Republic (TGM), the country in crisis (EB), rising discontent (AN), rapid changes and reform (LS), and the beginning of a democratic Czechoslovakia (VH)). Confirmation of this observation awaits further study.

values for each POS in speeches by each president. The DIN values in boldface are noteworthy (Table 4).

The table shows that socialist presidents noticeably prefer nominals (nouns and adjectives) and dis–prefer verbs, while others show the opposite tendency. In fact, DIN values for adjectives and verbs are in strong negative correlation (rho = −0.945): the lower the prominence level for adjectives, the higher the prominence of verbs. The prominence of adjectives suggests a contrastive and evaluative nature of text. The following example[21] (1) contrasts capitalism and socialism, and evaluates them by pointing out the shortcomings of the former:

(1)   **Kapitalistické** země přesunují vzestup **světových** cen surovin a materiálů do hladiny **spotřebitelských** cen a břemeno **surovinové** a **energetické** krize na bedra pracujících. To podporuje **prudkou** inflaci, nezaměstnanost a nejistotu pracujících. Touto cestou naše **socialistické** zřízení nemůže jít. 'Capitalist countries shift the rise of **world** prices for natural resources and materials onto the level of **consumer** prices and [shift] the burden of **natural resource and energy** crisis onto the workers' shoulders. This promotes **raging** inflation, unemployment and lack of security for workers. Our **socialist** system cannot go along this path.'
      (GH n1975)

The POS data above points to differences in the way speakers represent situations. "Noun–heavy" discourse points to the possibly frequent nominalization of actions, one of the major strategies in political texts to de–accentuate their components (Fairclough 2003: 144).

Another important property concerns the speaker: the image of the president as the leader of the state, as an active participant in situations, and as an individual who takes personal responsibility for his words and actions. The subsequent section will thus pursue the following two aspects of discourse[22]:
a.   decomposability of situations
b.   image of the speaker

---

21 Cf. KWA in Fidler and Cvrček (2015), which shows that adjectives related to socialist ideology (inflected word forms of adjectives *socialistický* and *komunistický*) are prominent in Gustáv Husák's speeches.
22 These two aspects of discourse are consistent with the important points of focus in media text analysis: "representations, identities, and relations" (Fairclough 1995: 5–12). Decomposability is connected with representation of the world (e. g., events) and with the components set up in the story (identities). Any image of the speaker is connected with identity (how the speaker is represented in the story) and the relations between the speaker and the other possible participants in events.

Table 4: DIN values of POS by presidents.

DIN	TGM	EB	EH	KG	AZ	AN	LS	GH	VH	VK	MZ
Nouns	0.59	−2.69	−1.07	**3.94**	**9.47**	**2.82**	**4.86**	**12.01**	−9.49	−5.58	−2.22
Adjectives	−3.71	−1.50	2.20	**4.72**	**7.35**	−1.96	**2.63**	**11.89**	−5.63	−1.95	−9.34
Pronouns	−6.73	**2.97**	**7.21**	−11.58	−10.54	−5.82	−0.25	−14.64	**9.48**	**6.42**	2.41
Numerals	−37.25	**5.86**	−22.28	**23.93**	−13.49	**13.10**	−13.49	3.44	−14.81	**8.64**	**22.38**
Verbs	**5.48**	−1.67	0.96	−5.80	−3.73	**2.17**	−4.03	−12.67	**4.14**	**2.66**	**8.44**
Adverbs	−2.62	**6.26**	−10.60	3.00	−11.89	−1.71	−5.09	−22.88	**9.31**	1.19	−4.36
Prepositions	−6.21	**1.85**	2.17	**5.35**	−0.62	**5.61**	0.88	**2.46**	−5.61	0.30	−8.10
Conjunctions	**3.37**	**3.76**	−7.53	−3.40	−0.18	−6.33	−3.32	−7.58	**4.00**	1.02	0.92
Particles	1.36	**14.95**	−29.54	−1.84	−26.59	−18.92	−12.47	−78.42	**18.71**	**9.49**	−5.77
Interjections	**62.77**	−100.0	78.92	**79.15**	−100.0	−100.0	−100.0	−100.0	−100.0	−100.0	−100.0

Boldface: noteworthy DIN value; RefC: SPEECHES.

Decomposability refers to an inherent ability of a linguistic form to provide specific information about the event (e. g., participants, tense, and modality). For example, the active voice finite form of the verb *poškodit* 'damage' requires the subject in the nominative case and the object in the accusative case: *Zloděj nejprve* **poškodil** *zámek dveří auta* [The thief **damaged** the door lock of the car] (SYN2015). The finite verb specifies who was responsible for the act of damaging and what was damaged; the indicative mood and perfective aspect of the verb form also explicitly report that the event did take place at a specific point in time. In contrast, the corresponding nominal form does not require the explicit presence of the participants: e. g., *v operačním systému došlo k* **poškození***, které brání správné funkci zvukového ovladače* [(lit) in the operational system it resulted in **damage**, which prevents correct functioning of volume control] (SYN2015); the agent of the damage is not obligatory, allowing the use of the impersonal construction *došlo k* [it resulted in] and reporting only the symptom (the malfunctioning of the volume control) rather than what caused the symptom.

Image of the speaker consists of two elements. One of them is the agency[23]; an element "of critical importance to critical discourse analysis" according to van Leeuwen (2008: 23) that is attributed to the speaker (e. g. whether the speaker presents a view as his own, or conflate his view with others).[24] The other element is the degree of explicit speaker participation in situations (e. g., to what extent the speaker is visibly represented as playing various roles).[25]

The following sections discuss explicit grammatical marking of participant[26] roles in narrated events (nouns vs. verb finite forms, case marking of nouns) to examine decomposability. We then look at verbal negation and the use of 1sg and 1pl grammatical marking to examine speaker image.[27]

---

23 The grammatical category of animacy in Czech, was not considered as an indicator of agency for two reasons: Animacy is relevant to masculine nouns only; moreover, animate nouns do not necessarily refer to human referents (e,g, *kůň* [horse], *virus* [virus]).

24 Agency is critical especially in presidential speeches, as leadership is one of the most important functions of the head of the state.

25 These two aspects of discourse are consistent with the important points of focus in media text analysis: "representations, identities, and relations" (Fairclough 1995: 5–12). Decomposability is connected with representation of the world (e. g., events) and with the components set up in the story (identities). Any image of the speaker is connected with identity (how the speaker is represented in the story) and the relations between the speaker and the other possible participants in events.

26 We use the term "participants" of the narrated event (Jakobson 1990[1957]) rather than social actors (van Leeuwen 2008: 7–8) to refer to entities and individuals that are *represented* as being responsible for actions in text, as the latter implies the *actual* individuals who are responsible for actions but may not be represented in text.

27 Here we obviously do not take into account the nature of political systems during different presidencies. Our discussion is purely based on morphosyntactic features and what they could

6 Decomposability

Narrated events with explicitly represented participants are decomposable; the audience has access to who or what is involved in the events. The following subsections will examine the decomposability of narrated events in verbal and nominal forms – verb finite forms vs. nouns and case marking of nouns – as indicators of decomposability.

6.1 Finite verb forms vs. nouns

A finite verb form suggests explicit reference to the participants of an action. The form thereby provides the audience access points to agree, modify, or contest who/what is responsible for an action, and who/what is affected. A finite verb form also allows the audience to question the nature of an action because of the explicit markings of tense, aspect, and modality. Compare the following texts with the verb finite form '[we] develop' and the corresponding de–verbal noun 'development':

(2)   Zajistit pokojný **rozvoj** a bezpečnost pro naši zemi.
        'To ensure calm **development** and safety for our country.'
        (LS n1970)

(3)   [...] upevňujeme demokratický politický řád a **rozvíjíme** naši mladou tržní ekonomiku.
        '[...] [we] consolidate the democratic political order and **develop** our young market economy.'
        (VH n2003)

    In (2) the existence of the action ("development") is taken for granted; what is important is that the action be calm and uneventful. The nominal phrase does not provide specific aspects of the event; i. e., who carries out the action and what is being developed. In (3) the addressee could potentially contest or modify the subject (Who are "we"? The political elite or the whole nation?), and/or the predicate (To what extent are we developing the market economy? How are we executing it?).

---

potentially indicate. It will be clear later on that the nature of presidency is consistent with the type of discourse, which morphosyntactic features indicate.

The results below split the presidential speeches into roughly two groups: +/–the 'noun–heavy' speakers. The values in boldface are deemed noteworthy. The DIN values for verbs (POS) are reproduced for comparison to the verb finite forms (Table 5).

Table 5: DIN values of verb finite forms and nouns.

DIN	TGM	EB	EH	KG	AZ	AN	LS	GH	VH	VK	MZ
Verb fin. (VF)	**6.51**	−1.25	1.90	−3.83	−6.15	**2.27**	−3.92	−9.49	1.44	**3.18**	**8.44**
Verbs (POS)	**5.48**	−1.67	0.96	−5.80	−3.73	**2.17**	−4.03	−12.67	**4.14**	**2.66**	**8.44**
N (POS)	0.59	−2.69	−1.07	**3.94**	**9.47**	**2.82**	**4.86**	**12.01**	−9.49	−5.58	−2.22

Boldface: noteworthy DIN value; RefC: SPEECHES.

The table shows a major difference between most of the socialist presidents' speeches and the others. The former are noun–heavy, while the latter tend to use nouns below average while preferring verbs (at the expected frequency or above). By excluding non–finite forms from the verbs, we can state with increased confidence that most of the socialist presidents' speeches are less decomposable than the others. The next section on nominal case marking will further delineate discourse strategies of individual speakers.

6.2 Grammatical case–marking (nom, dat, acc, gen, and instr)[28]

Case endings mark syntactic roles of entities, thereby marking different participant roles in narrated events. The description of case is not only based on case semantics in traditional Czech grammar (Rusínová and Nekula 1995), but also is anchored in a cognitive linguistic view that case is meaning–bearing. Thus, each case marking represents consistent participant roles and positions in the event; these meanings are instantiations of a more abstract semantic schemas (Janda 1993; Janda and Clancy 2006).

---

28 Czech nominal declension also includes the locative and the vocative cases. These cases, however, are not relevant to the discussion in this section; the locative is always used in prepositional phrases (therefore deemed to play a less central role in events), and the vocative case was excluded from the discussion because of its genre–specific property discussed in Section 2.3. The data in this section concern only full nouns and do not include adjectives or pronouns.

The nominative is likely to refer to the responsible entity in most sentences:

(4)   [...] se **někteří Poláci**[nom] snaží zmocnit našeho území, [...]
       '[...] **some Poles**[nom] try to usurp our territory [...]'
       (TGM v1918)

On the contrary, the accusative is most likely to refer to the patient of an action (Rusínová and Nekula 1995: 240) or destination or dimension where force is applied (Janda and Clancy 2006: *111–145*).

(5)   Zkrotit a isolovat **dobrodruhy**[acc] z tábora imperialistických útočníků, [...]
       'Taming and isolating **the adventurers**[acc] from the camp of imperialist attackers, [...]'
       (KG n1953)

The other cases have functions that are not connected specifically to the agent or patient. The main function of the instrumental case is said to be adverbial, indicating place, itinerary, time, means, reason, circumstances, and the respect to which some statement is valid (Rusínová and Nekula 1995: 242), or "accessary for something else" (Janda and Clancy 2006: 181).[29] In SPEECHES the instrumental is most strongly collocated with the preposition *s* 'with' to denote accompaniment to an action, which serves as an adverbial phrase.

(6)   **S dnešní otevřeností**[instr] světa žít musíme, [...]
       '[We] must live **with today's openness**[instr], [...]'
       (MZ r2013)

(7)   Odešel do exilu **se svým snem**[instr], **se snem**[instr] samostatnosti Československa.
       '[He] emigrated **with his dream**[instr], **with** [a] **dream**[instr] of independence of Czechoslovakia.'
       (MZ r2013)

The dative case is associated with multiple functions, e. g., the roles of beneficiary, experiencer, or victim (Rusínová and Nekula 1995: 239).[30] In

---

29 The instrumental case in Czech can have an adverbial function both in the active and the passive voice. In the passive voice the instrumental can in principle function as a demoted subject with reduced agency, but this use is not predominant in SPEECHES.
30 Janda and Clancy also show the meaning of a competitor (2006: 61–62).

SPEECHES the dative is strongly collocated with two prepositions: *k* 'towards' and *proti* 'against', and is associated with references to purpose and conflict (implying the existence of opponent(s)), as in the following examples:

(8) Příchod revoluční vlády soudruha Kádára a pomoc Sovětského svazu této vládě v boji **proti kontrarevolučním bandám**^{dat} [...] (AZ n1957)
     'The arrival of comrade Kádár's revolutionary government and the assistance of the Soviet Union to this government in [its] battle **against the counter–revolutionary gangs**^{dat} [...]'

(9) Zachováme–li nadále bezpodmínečnou jednotu a národní kázeň, zaručíme **národu**^{dat} i každý sobě šťastnou budoucnost.
     'If [we] preserve further the unconditional unity and the national discipline, [we] will guarantee **the nation**^{dat} and everyone [will guarantee] oneself a happy future.'
     (EH v1943)

The genitive case functions mainly as an attribute (Rusínová and Nekula 1995: 237), as in the following examples:

(10) V souladu s vůlí a životními zájmy **československého lidu**^{gen} podporujeme mírovou politiku **Sovětského svazu**^{gen} a aktivně se podílíme na jejím prosazování.
     'In agreement with the will and the vital interests **of Czechoslovak people**^{gen} [we] support peaceful policy **of the Soviet Union**^{gen} and actively take part in its support.'
     (GH n1983)

In the example above, the genitive case 'of the Soviet Union' specifies that Czechoslovakia supports not just any peace–making politics, but specifically those of the USSR, implicitly referring to its counterpart (e. g. the "alleged" peace–making but "imperialist" US politics). The action is in the interest of not just any nation, but in the interest *of the Czechoslovaks*.

It is also important to note that the genitive case can blur the differences between two diametrically opposite participant roles – the semantic subject and the object: the same form can represent the agent (*zneužít nespokojenosti **mas*** [to abuse the discontent **of the masses**] (AZ n1957)) and the patient (*překonávání všech **nedostatků*** (GH n1975) [overcoming of all the **deficiencies**']). The genitive case can be considered as another morphosyntactic device to avoid explicit marking of who is responsible for the action; this function is consistent with

the observation in Janda and Clancy who state that the genitive brings an entity to the background (2006: 24).

Table 6 below shows the DIN value for each case form as used by each president. The numbers in boldface are deemed noteworthy. The DIN values for the POS are reproduced at the bottom for comparison.

Table 6: DIN values of grammatical cases.

DIN	TGM	EB	EH	KG	AZ	AN	LS	GH	VH	VK	MZ
Nom.	**13.95**	**2.88**	−0.84	0.71	0.89	−2.59	−3.00	−5.23	−3.94	−5.24	**8.62**
Instr.	**4.96**	−2.2	−1.92	−1.29	7.25	−6.28	5.73	8.87	−7.99	3.16	5.88
Dat.	4.01	−7.19	**14.53**	−6.38	**16.92**	2.33	3.15	3.66	−3.34	−3.77	0.39
Gen.	−6.22	−6.29	−4.09	**8.25**	**15.32**	5.59	8.84	**21.56**	−15.13	−9.27	−13.41
Acc.	−0.8	−5.67	−0.05	2.85	**9.5**	1.58	4.57	9.96	−4.78	−2.74	−0.01
Nouns (POS)	0.59	−2.69	−1.07	**3.94**	**9.47**	**2.82**	**4.86**	**12.01**	−9.49	−5.58	−2.22

Boldface: noteworthy DIN value; RefC: SPEECHES.

The table shows no significant divergence from the expected frequency for the nominative in the socialist presidents' texts. Instead, their DIN values for the genitive are mostly the highest compared to the other cases. The tendency to overuse the genitive, which often functions as an attribute, is commensurate with the adjective–predominant nature of socialist presidents' speeches[31] (Table 4). It is also notable that the socialist presidents use the genitive, which is likely to obfuscate the agent and the patient. The pattern is particularly striking when we consider that the genitive is a core constituent within the entire SPEECHES (Table 1); the socialist presidents' use of the genitive is highly noticeable within the already genitive–heavy corpus.

In contrast, the democratic presidents do not show positive DIN values for the genitive. More than half of the democratic presidents' speeches show noteworthy DIN values for the nominative case (TGM, EB, MZ), and the instrumental case (TGM and MZ). Their use of the nominative is striking, especially as these are not noun–heavy presidents (Table 4) and the nominal case and the POS in general are strongly correlated (Table 2). The high score for the nominative suggests the importance of verb phrases in discourse for which the

---

[31] Admittedly, adjectives can be both attributive and predicative. However, adjectives are predominantly used as attributes in the socialist presidents' speeches (530,749.65 i.p.m. adjective preceding noun out of 713,739.64 i.p.m. for all adjectives = 74.4 %).

nominative serves as the grammatical subject. The prominent use of the instrumental case (TGM) suggests that predicates are often further elaborated and are decomposable (e. g. explaining *how* actions are carried out). The table further confirms that the democratic presidents' texts are more accessible to the audience for questioning and contesting (decomposable) than those of the socialist presidents.[32] Among the democratic presidents, TGM, which show prominence in both the nominative and the instrumental, appear to be most decomposable and therefore most accessible to the audience to engage in discussion.

EH, the Czechoslovak president under the Nazi occupation, stands out; only the dative is noteworthy in his speeches, while all other cases show expected levels of frequency (i. e., they are statistically insignificant). The prominence of the dative suggests that EH's texts particularly focus on goals and purposes as well as potential opponents rather than reporting who is responsible for actions or how actions are carried out, as in the following example.[33]

(11) [...], že stejně jako přijde každá drobná zásluha jednotlivcova **k dobru**[dat]
       celému národu[dat], bylo by i každé vybočení jednotlivcovo velikým
       ohrožením celku.
       '[...] that there comes every small service of an individual **to the good**[dat]
       for the entire nation[dat] in just the same way that every divergence by an
       individual would be a threat to the whole.'
       (EH 1943)

Prominence of the dative case such as (11) suggests that the speaker is not empowered[34]; actions are justified not because of the speaker's visions, but because of threats and/or goals (typically national goals).

---

32 VH's texts show negative DINs for all cases and VK's texts the absence of noteworthy DINs for all case forms. The results should be seen in connection (a) with their DIN values for the overall POS (noun), which are the lowest among all the presidents, and (b) with the strong correlation between cases and nouns (see Table 2). The numbers in Table 6 are likely to have resulted from the VH's and VK's scarce use of nouns in general.

33 Specific use of the dative case in EH's speeches is also observable among personal pronouns. EH's DIN value for the dative form *nám* [to us] of the personal pronoun *my* [we] is the highest among all the presidents.

34 The dative case is said to express "the referent's lowered responsibility and lack of control over the event encoded by the predicate" also in Russian (Perelmutter 2010: 3218; Israeli 1997; Timberlake 2004).

6.3 Decomposability: Discussion

Section 6 examined how narrated events are represented, for which the notion of decomposability was used as an indicator: the more decomposable a situation is represented, the more access points are available for the audience to engage with the representation of situations. The verbal and nominal features in this section confirmed and elaborated differences between the socialist and democratic presidents' texts. Moving from a KMA based on POS to a KMA based on lower–level morphosyntactic features, we obtained a sharper contrast between these two groups.

The socialist presidents' speeches provide limited access for the audience to question the different components in the narrated events; attributive function is predominant in their speeches. By contrast, the democratic presidents' speeches tend to be decomposable and provide points of discussion for the addressee. While the socialist presidents report "what something is like"/"how something is", their democratic counterparts report more about "who does what and how".

The results from this section also highlight the discourse properties of some individual speakers. TGM's texts show high DIN values for both the nominative and the instrumental cases, suggesting that the texts represent situations in such a way that they are most accessible for the audience to question, contest, and challenge the speaker's views. The data in this section also contrast EH's discourse from all the other speakers. This pattern is perhaps to be expected because of his position as president under occupation, but KMA shows that morphosyntactic information can provide purely textual empirical evidence for the speaker's position, which forced him to emphasize goals and threats to justify situations and actions rather than referring to his own personal responsibility.

The following section will explore the representation of the speaker as an agent and as a visible participant. More individual differences will emerge.

7 Image of the speaker

Discussion of speaker image is based on several morphosyntactic features: verbal negation,[35] the use of the 1sg and 1pl indicative present tense (ind–pres) verb forms, and the 1 sg and 1pl personal and possessive pronouns (pers–prn, pos–prn)

---

35 Verbal negation is not syntactic in Czech, but it is part of a conjugated verb form (e. g. *nejsem* [[I] am not] vs. *jsem* [[I] am]).

The 1sg and 1pl ind–pres forms are lower–level morphosyntactic features with respect to verbs, and 1sg and 1pl pers–prn, pos–prn forms with respect to pronouns (Table 2).[36]

7.1 Implicit agency: Verbal negation[37]

Verbal negation reflects degrees of implicit agency; the speaker implicitly considers and corrects (and therefore acts upon) the addressee's misguided belief.[38] This feature can also attribute to the speaker authority over the audience, i. e., an image that the speaker "knows better" than the audience.[39] Table 7 shows the DIN values for verbal negation for each president. The DIN values for verbs in general are reproduced at the bottom for reference.

Table 7: DIN values of verbal negation.

	TGM	EB	EH	KG	AZ	AN	LS	GH	VH	VK	MZ
Verbal negation	**16.93**	−6.46	0.04	−23.51	−4.47	−11.19	−17.80	−61.17	**9.73**	**28.78**	1.31
Verbs (POS)	**5.48**	−1.67	0.96	−5.80	−3.73	**2.17**	−4.03	−12.67	**4.14**	**2.66**	**8.44**

Boldface: noteworthy DIN value; RefC: SPEECHES.

Verbal negation is noteworthy in texts by TGM, VH, and VK, e. g.:

(12) **Nežijeme** v zemi, kde jsou demonstranty zapalovány tisíce aut a demolovány stovky výloh obchodů. **Nežijeme** v zemi, kde narůstají problémy s

---

36 Besides the 1sg pers–prns, which are labeled in Table 1 as "common" (Class 3), all the other features examined in this section belong to the "frequent features" (Class 4). Both groups are numerous enough to be used in quantitative analysis.

37 By negation we mean negated verbs (negation is attached to a conjugated verb form in Czech). Negation in other parts of speeches (esp. adjectives and adverbs) tends to be lexicalized (Kováříková et al. 2012).

38 Ueda (1992) shows that such a discourse function involving negation can motivate the case–marking variation in the Russian predicate adjectives.

39 If verbal negation is extremely prominent, the text may therefore project an image of a speaker who is better informed, knowledgeable than the audience. This could even be interpreted as expression of arrogance in some situations. It is not surprising that such a negative assessment is sometimes attributed to VK, whose DIN value for verbal negation is distinctly high in this table.

množstvím přistěhovalců, kteří svým stylem života narušují základní soudržnost země. **Nekrachují** u nás banky, stát je **nemusí** zachraňovat za peníze nás všech. Naše míra nezaměstnanosti je pod průměrem zemí Evropské unie.

'[We] are **not living** in a country where thousands of cars are burnt by demonstrators and hundreds of shop windows are demolished. [We] are **not living** in a country where there are growing problems with multitudes of immigrants who disturb the basic cohesiveness of the country with their lifestyle. Our banks are **not going broke**; the state **does not have to** save them at the expense of all our money. Our unemployment rate is under the EU average.'
(VK n2012)

Example (12) can be seen as a reaction to the existing false assumption that the conditions in the Czech Republic are terrible.

Verbal negation is not prominent in any of the socialist presidents' speeches. Lack of such implicit interaction with the addressees is consistent with their discourse style that does not easily allow the addressees to engage in discussion (Section 6).

7.2 Explicit agency of the speaker and the speaker as a visible participant in text: 1sg forms

The 1sg grammatical marking specifically points to the speaker. Morphosyntactic forms with this marking, however, project different types of speaker image. The 1sg ind–pres verb forms suggest that the speaker takes the responsibility or claims credit for his/her own actions and statements, as in the following example, where there is no question about the agency of actions.

(13) Mnoho z vyznamenaných muselo čelit nejrůznějším druhům závistivých pomluv. Pomluv těch, kterým **říkám**[1sg–ind] komentátoři.
'Many of the awardees had to face the most various types of envious slanders. Slanders from those [I] **call** commentators.'
(MZ r2013)

Prominence for the 1sg pers–prns indicates that the speaker is represented as a visible participant in the text; the speaker plays different roles in text by appearing in various syntactic roles.

(14) Tyto diskuze jsou jistě plodné, ale něco **mně**[1sg–pers–prn–dat, emphatic] v nich chybí, chybí **mi**[1sg pers–prn–dat] v nich jisté odosobnění českých dějin.
'These discussions are certainly productive, but *to me* something is missing in them, certain depersonalization of Czech history is missing in them **to me** [to my mind].'
(MZ v2013)

The speaker in (14) is not the agent, but it is the experiencer, one who feels "something is missing".

A visible participant role in text may occasionally overlap with the role of an agent when the pres–prn appears in the nominative case. Use of the pronoun in the nominative to express agency, however, is marked in Czech; the form is used only exclusively for an emphatic–contrastive effect in formal style:

(15) **Já**[1sg– pers–prn–nom, emphatic] naopak myslím, že Lenin je Rus a typický Rus, [...]
'*I* on the contrary think that Lenin is a Russian, and a typical Russian, [...]'
(TGM r1919)

The speaker in (15) makes an implicit contrast between his opinion and the opposing view, which is likely to be entertained by others.

Unlike the 1sg ind–pres verb forms and 1sg pers–prn forms, the 1sg poss–prn forms assume an attributive function. They report that entities belong personally to the speaker:

(16) **Můj**[1sg–pos–prn] vztah k Slovensku je znám.
'**My** [personal] relationship to Slovakia is known.'
(LS n1969)

Besides reporting the possessor, the 1sg pers–prn can report the embedded subject of nominalized actions:

(17) **Moje**[1sg–pos–prn] dnešní cesta do Moskvy
'**My** trip today to Moscow'
(EB v1943)

As part of the nominalized expression, the pos–prns used in this context can suppress the agency of the speaker (as seen in section 3.2.). The agent is not represented as controlling different phases of the event. This nominalized phrase does not report modality: if the speaker feels that it is his duty to carry out the action, or whether he is forced to carry it out, allowed to carry it out, or

voluntarily carries it out. Clearly, the 1sg pos–prn, compared to 1sg ind–pres verb forms and 1sg pers–prns, attributes weaker agency to the speaker.

Table 8 shows the results obtained with respect to the 1sg forms. The DIN values in boldface are noteworthy. The DIN values for verbs are reproduced for comparison. The DIN values for pronouns are not compared with the 1sg pronominal forms, as neither one of them is very strongly correlated with the part of speech (Table 2); any positive DIN values that are significant can be considered noteworthy.

Table 8: DIN values of 1sg forms.

	TGM	EB	EH	KG	AZ	AN	LS	GH	VH	VK	MZ
1.sg. ind pres.	**36.78**	1.68	12.64	−72.53	−44.34	−17.81	−10.50	−30.15	**7.72**	−0.66	**24.26**
Verbs (POS)	**5.48**	−1.67	0.96	−5.80	−3.73	**2.17**	−4.03	−12.67	**4.14**	**2.66**	**8.44**
1. sg pers pron	**24.70**	1.28	**30.00**	−46.88	−73.57	−22.83	−7.20	−78.51	2.45	−3.77	**75.87**
1. sg poss pron	20.51	11.06	**70.95**	−36.02	−100.00	−69.83	16.92	−76.91	5.65	−24.75	**58.87**

Boldface: noteworthy DIN value; RefC: SPEECHES.

Table 8 above shows a distinct property of the socialist presidents' speeches; none of them shows a noteworthy DIN value.[40] The table also suggests individual differences in discourse among the non–socialist presidents.

High degrees of explicit speaker agency are found in texts by TGM, VH and MZ with some differences. In VH's texts the speaker is represented as the agent of actions and statements (1sg ind–pres), but not in a variety of participant roles (1sg pers–prn). On the contrary, the speaker is highly visible also in various participant roles in speeches by TGM and MZ. MZ differs from both TGM and VH in showing high DIN values for all the three forms. MZ, then, seems to manipulate agency, emphasizing the speaker's role as the agent and active participants, but noticeably obscuring the speaker agency when it is necessary.

VK and EB do not express explicit agency or speaker visibility, nor do they suppress agency by using pos–prns. VK, however, differs from EB in the prominent use of verbal negation (Table 7); VK's speeches do not outwardly express agency, but highlight the speaker's agency by pitting his opinions against the audience's. EB is the only speaker among the non–socialist presidents whose DIN values do not significantly depart from the expected values.

---

40 The results may be partially connected with the generally low DINs of verbs and pronouns.

The DIN values also suggest that EH's speeches again depart from the others. EH's texts render the speaker less agentive than the other non–socialist presidents.

His DIN values show neither explicit nor implicit agency (insignificant DIN value for the 1sg ind–pres forms and for verbal negation). However, EH's speeches present the speaker as a visible participant in situations (1sg pres–prns) and possessor or suppressed agent in nominalizations (1sg pos–prns), as in the following example:

(18) [...] chtěl bych o svátcích míru a radosti říci Vám po prvé několik slov s místa, na něž **mne**[1sg-pers-prn-acc] postavila ústavně projevená vůle národa a **moje**[1sg-pos-prn] vlastní povinnost.

'During these holidays of peace and joy I would like to tell you first of all some words from the position, onto which the constitutionally expressed will of the nation and **my** own obligation placed **me**.'
(EH v1938)

7.3 Opacity of speaker agency and visibility: 1pl forms

Unlike 1sg person marking, the grammatical marking of the 1st pers plural inherently reports opaque referentiality. Referents of the 1pl forms include the speaker but also simultaneously refer to others without specifying their identity: the entire audience who are listening to the speech, the entire nation (including those who may not be among the audience), only some of the audience selected by the speaker, or a group of people with whom the speaker is affiliated (e. g., the Central Committee of the Communist Party).[41] Observe the following verb forms in boldface:

(19) Vykonanou práci přitom **hodnotíme**[1pl-ind] střízlivě. **Víme**[1pl-ind], že některé kolektivy vstupují do nového roku s nesplněnými úkoly. Otevřeně **kritizujeme**[1pl-ind] slabiny a nedostatky, které se na různých místech stále vyskytují.

'At the same time [we] soberly **evaluate** the work done. [We] **know** that some collectives enter the new year with unfulfilled tasks. [We] openly

---

41 The 1pl grammatical marking can even refer specifically to the speaker in other contexts: the Royal *we* or the author's *we*. Whether the form marks solidarization with the addressee is not relevant here; what is relevant is that the form is inherently opaque as to who shares the opinions or carries out actions with the speaker.

criticize the weaknesses and shortcomings, which still occur in various places.'
(GH n1986)

Those who evaluate the work, those who know, and those who criticize may be the speaker and some of the select members of the party; alternatively, everyone may be expected to evaluate, know, and criticize.

When the 1pl pronominal forms are prominent in a text, it is likely that the text obfuscates who exactly participates in situations, and who exactly is the possessor or suppressed agent. The 1pl pronominal forms below may refer to the speaker and the audience, the entire nation, or the political leadership:

(20) Nenahraditelný význam pro **nás**[1pl–pers–prn–acc] mají zkušenosti Sovětského svazu a jiných bratrských socialistických zemí, [...].
'The experience [lit. experiences] of the Soviet Union and other brotherly socialist countries, [...] have an irreplaceable significance for **us**.'
(GH n1988)

Finally, nominal phrases with the 1pl pos–prn 'our' in (21) below may refer to the entire nation, the party leadership (including the speaker), or only part of the nation that agrees with the speaker and the current political order (i.e. excluding those who are sabotaging).

(21) **Naše**[1pl–pos–prn] výstavba socialismu je již takovou mohutnou silou, že přes všechny škody a sabotáže, které spiklenečtí zrádci napáchali, nepodařilo se jim rozvrátit a zlomit vývojovou linii **našeho**[1pl–pos–prn] hospodářství.
'**Our** construction of socialism is already such a massive power that despite all the damages and sabotages, which the conspiring traitors committed, they did not succeed in overturning and breaking the developing line of **our** economy.'
(KG n1953)

The nominalized action below (*náš vývoj* [our progression]) contains an embedded agent ("we"):

(22) **náš**[1pl–pos–prn] vývoj k vyspělým národům s rozvinutou ekonomikou, bohatým společenským i kulturním životem
'**our** progression towards matured nations with a developed economy, rich social as well as cultural life'
(LS r1968)

Again, the pronoun *náš* 'our' may refer to the entire nation, or to only that part of the population that participates in its progression (i.e., those who agree with the regime and work with the system), or those who lead the progression (the political elite).

The pattern in Table 9 below differs from that in Table 8.

Table 9: DIN values of 1pl forms.

	TGM	EB	EH	KG	AZ	AN	LS	GH	VH	VK	MZ
1pl ind pres	−5.07	−27.60	11.69	−15.76	−5.32	**10.80**	8.28	**15.13**	3.35	**8.16**	−33.35
Verbs	**5.48**	−1.67	0.96	−5.80	−3.73	**2.17**	−4.03	−12.67	**4.14**	2.66	**8.44**
1pl pers pron	4.16	**9.78**	**15.01**	−30.09	−8.19	−9.17	−9.25	−35.58	3.88	**21.22**	−61.20
1pl poss pron	−1.37	−7.62	8.12	**14.91**	4.72	3.25	**15.18**	**14.65**	−12.66	−0.04	−57.87
Pronouns	−6.73	**2.97**	**7.21**	−11.58	−10.54	−5.82	−0.25	−14.64	**9.48**	**6.42**	2.41

Boldface: noteworthy DIN value; RefC: SPEECHES.

The table shows that use of opaque referentiality is found in speeches by presidents, regardless of the ideological divide (EB, EH, KG, AN, LS, GH, VK). However, it is worth noting that the majority of the socialist presidents, whose DIN values for the POS are predominantly negative (or neutral), show prominence in at least one of the opacity features. Among these speakers, GH is particularly striking; in spite of his below−average use of pronouns and verbs (Table 2), GH's use of 1pl. ind. and 1pl. pos−prn forms are unexpectedly high. His texts are highly marked with opacity for agency and support the hypothesis in our previous study (Fidler and Cvrček 2015) that pos−prn forms are important keys to the interpretation of GH speeches. Strong opaque agency is consistent with the period of political stagnation (1970s and 1980s).

Some comments are necessary regarding the democratic presidents who do not show statistically significant DIN values in any form in Table 9: TGM, VH, and MZ. Based on the results from Tables 8 and 9, it is reasonable to state that their speeches project an image of a speaker as markedly agentive (TGM, VH, MZ). Differences among them can be nonetheless found in Tables 7 and 8, as discussed in the preceding sections.

8 Conclusions

KMA probes aspects of text that KWA does not address. While KWA focuses on lexicon and provides a key to genre, topics, and semantic associations in text,

KMA focuses on morphosyntactic features with a minimum amount of lexical information and provides a key to more general properties of discourse; it informs the way situations and their participants are consistently represented. We showed how KMA helps reveal two general aspects of discourse: representation of situations and speaker image in presidential speeches. KMA reflects discourse properties symptomatic of distinct political systems, discourse properties shared by speakers from different political systems, and discourse properties specific to individual speakers. By doing so, this study essentially demonstrates that the concept of keyness (introduced in KWA) can be extended to a more abstract level beyond specific textual content. While KWA reveals what the text is about, KMA in this paper reveals how a text allows the addressee to confirm or question situations and statements represented in text, how the speaker is visible as a participant in situations, and how the speaker takes responsibility for statements and situations.

This study also yielded some offshoots. We showed a new concept that is necessary for KMA (noteworthiness) because of the specific hierarchical nature specific to keymorphs. We also showed that keyness of POS is more informative than previously found by Culpeper (2009); POS, as part of KMA, is informative for examining discourse strategies rather than topics.

KMA in this paper led to some specific findings about the actual speeches. Higher–level morphosyntactic features (POS or a more widespread morphosyntactic feature such as finite verb forms) point to a rough ideological divide among speakers; KMA based on higher–level morphosyntactic features captures discourse characteristic of socialist presidents' speeches as attributive and noun–heavy. Lower–level morphosyntactic features (case, grammatical person) inform individual speakers' discourse more than their higher–level counterparts: KMA empirically confirmed the differences between TGM and VH, who were highly respected statesmen, and the other democratic presidents. The speeches by TGM and VH are not only less noun–heavy, but also interactive and consistently agentive.

On a more general level and for further research, the paper indicates that the hierarchical nature of morphosyntactic features parallels the hierarchical nature of discourse properties: the higher, ideological differences can be probed with higher–level morphosyntactic features, and individual speaker differences with lower–level morphosyntactic features. Lower–level morphosyntactic features promise to be applicable to future research on the manipulation of agency and accountability in discourse.

This study does not claim to provide a full picture of the discourse properties of each speaker. The list of the morphosyntactic features we considered is by no means exhaustive. Some of the observations on individual speakers await

further inquiry. As KWA and KMA yield different types of information about discourse, synthesis of keyness[42] from different levels – both lexical and morphosyntactic information – should be considered for a comprehensive corpus–assisted discourse analysis. We attempt to do so in a separate study.

References

Altshuler, Daniel. 2010. Aspect in English and Russian flashback discourses. *Oslo Studies in Language* 2. 75–107.

Baker, Paul. 2009. The question is, how cruel is it? Keywords in debates on fox hunting in the British House of Commons. In Dawn Archer (ed.), *What's in a word–list?* 125–136. London: Ashgate.

Biber, Douglas. 1993. Using register–diversified corpora for general language studies. *Computational Linguistics* 19(2). 219–241.

Biber, Douglas. 2006. *University language: A corpus–based study of spoken and written registers*. Amsterdam: John Benjamins.

Chvany, Catherine. 1990. Verbal aspect, discourse saliency, and the so–called perfect of result in Modern Russian. In Nils B. Thelin (ed.), *Verbal aspect in discourse*, 213–236. Amsterdam: John Benjamins.

Culpeper, Jonathan. 2002. Computers, language and characterisation: An analysis of six characters in *Romeo and Juliet*. In Ulla Melander–Marttala, Carin Ostman & Meria Kyto (eds.), *Conversation in life and in literature: Papers from the ASLA symposium, Association Suédoise de Linguistique Appliquée (ASLA) 15*, 11–30. Uppsala: Universitetstryckeriet.

Culpeper, Jonathan. 2009. Keyness: Words, parts–of–speech and semantic categories in the character–talk of Shakespeare's *Romeo and Juliet*. *International Journal of Corpus Linguistics* 14(1). 29–59.

Culpeper, Jonathan & Jane Demmen. 2015. Keywords. In Douglas Biber & Randi Reppen (eds.), *The Cambridge handbook of English corpus linguistics*, 90–105. Cambridge: Cambridge UP.

---

42 Fidler and Cvrček (forthcoming).

Cvrček, Václav, Petr Truneček & Horký Václav. 2015. *SPEECHES: korpus prezidentských projevů* [SPEECHES: Corpus of presidential speeches]. Prague: Ústav Českého národního korpusu FFUK. https://www.korpus.cz (accessed 1. March 2017).

Desclés, Jean–Pierre & Guentschéva Zlatka. 1990. Discourse analysis of aorist and imperfect in Bulgarian and French. In Nils B. Thelin (ed.), *Verbal aspect in discourse*, 237–261. Amsterdam: John Benjamins.

Divjak, Dagmar & Laura A. Janda. 2008. Ways of attenuating agency in Russian. *Transactions of the Philological Society* 106(2). 138–179.

Fairclough, Norman. 1995. *Media discourse*. London: Hodder Education.

Fairclough, Norman. 2003. *Analysing discourse: Textual analysis for social research*. London: Routledge.

Fidler, Masako & Cvrček Václav. 2015. A Data–driven analysis of reader viewpoints: Reconstructing the historical reader using keyword analysis. *Journal of Slavic Linguistics* 23(2).197–239.

Fidler, Masako & Cvrček Václav. Forthcoming. "Measuring" propaganda. In Masako Fidler & Václav Cvrček (eds.) *Taming the corpus: From inflection to lexis to interpretation*. New York: Springer.

Fielder, Grace. 1990. Narrative context and Russian aspect. In Nils B. Thelin (ed.), *Verbal aspect in discourse*, 263–284. Amsterdam: John Benjamins.

Fowler, Roger. 1991. *Language in the news. Discourse and ideology in the press*. London: Routledge.

Hopper, Paul & Sandra Thompason. 1980. Transitivity. *Language* 56(2). 251–299.

Israeli, Alina. 1997. *Semantics and pragmatics of the "reflexive" verbs in Russian*. Munich: Otto Sagner Verlag.

Jakobson, Roman. 1990[1957]. Shifters and verbal categories. In Linda R. Waugh & Monique Monville–Burston (eds.), *On language*, 386–392. Cambridge, MA: Harvard University Press.

Janda, Laura A. 1993. The shape of the indirect object in Central and Eastern Europe. *The Slavic and East European Journal* 37(4). 533–563.

Janda, Laura A. & Steven Clancy. 2006. *The case book for Czech*. Bloomington, IN: Slavica.

Kováříková, D., L. Chlumská & V. Cvrček. 2012. What belongs in a dictionary? The example of negation in Czech, *Euralex*. Oslo. http://www.euralex.org/elx_proceedings/Euralex2012/pp822–827%20Kovarikova,%20Chlumska%20and%20Cvrcek.pdf (accessed 1 March 2017).

Křen, Michal, Václav Cvrček, Tomáš Čapka, Anna Čermáková, Milena Hnátková, Lucie Chlumská, Tomáš Jelínek, Dominika Kováříková, Vladimír Petkevič, Pavel Procházka, Hana Skoumalová, Michal Škrabal, Petr Truneček, Pavel Vondřička & Adrian J. Zasina. 2016. SYN2015: Representative corpus of contemporary written Czech. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Sara Goggi, Marko Grobelnik, Bente Maegaard, Joseph Mariani, Hélène Mazo, Asunción Moreno, Jan Odijk & Stelios Piperidis (eds.), *Proceedings of the tenth international conference on language resources and evaluation (LREC'16)*, 2522–2528. Portorož: ELRA. http://www.lrec–conf.org/proceedings/lrec2016/pdf/186_Paper.pdf (accessed 1, March 2017).

Kresin, Susan. 1998. Deixis and thematic hierarchies in Russian narrative discourse, *Journal of Pragmatics* 30(4). 421–435.

Perelmutter, René. 2010. Impoliteness recycled: Subject ellipsis in Modern Russian complaint discourse. *Journal of Pragmatics* 42. 3214–3231.

Rusínová, Zdenka & Marek Nekula. 1995. Morfologie [Morphology]. In Petr Karlík, Marek Nekula & Zdenka Rusínová (eds.), *Příruční mluvnice češtiny [A Reference grammar of Czech]*, 227–358. Prague: Lidové noviny.

Scott, Mike. 2010. Problems in investigating keyness, or cleansing the undergrowth and marking out tails …. In Marina Bondi & Mike Scott (eds.), *Keynesss in texts*, 43–57. Amsterdam: Benjamins.

Scott, Mike & Christopher Tribble. 2006. *Textual patterns: Keyword and corpus analysis in language education*. Amsterdam: Benjamins.

Sonnenhauser, Barbara. 2008. Aspect interpretation in Russian – a pragmatic account. *Journal of Pragmatics* 40(12). 2077–2099.

Těšitelová, Marie. 1995. O jazyce prezidenta dr. Edvarda Beneše (1884–1948) [About the Language of Dr. Edvard Beneš], *Naše řeč [Our Speech]* 78(3). 124–135.

Timberlake, Alan. 2004. *A reference grammar of Russian*. Cambridge: Cambridge University Press.

Ueda, Masako. 1992. *The interaction between clause–level parameters and context in Russian morphosyntax: Genitive of negation and predicate adjectives*. Munich: Sagner.

Van Dijk, Teun. 1993. *Discourse and elite racism*. London: Sage.

Van Leeuven, Theo. 2008. *Discourse and practice: New tools for critical analysis*. Oxford: Oxford University Press.

Wodak, Ruth. 2009. *The discourse of politics in action. Politics as usual*. Basingstoke, UK, New York: Palgrave Macmillan.