

Calculating the Victory Chances

A Stylometric Insight into the 2018 Czech Presidential Election

Research Aims

- To prove the utility of the used tools in political-language research
- To provide a quantitative overview of the Czech 2018 presidential election
- To investigate stylistic differences among individual candidates
- To uncover marketing and rhetoric strategies used in their texts

Research Structure

- Stylometric assessment of the sample new-year speeches (QUITA)
 - 8 texts written on the request of Český rozhlas + Miloš Zeman's 2017 Christmas Address
- MFW analysis of the same (STYLO package in the R software)
- Contrastive analysis of two candidates' programmes (LancsBox)
 - Michal Horáček and Marek Hilšer
- Contrastive analysis of the final Czech TV debate (LancsBox)
 - Miloš Zeman and Jiří Drahoš

Election Overview

- Taking place: Jan 12–13; the second round: Jan 26–27
- 9 candidates (presented later)
- The first round: Miloš Zeman – 38.56%; Jiří Drahoš – 26.6%
- The second round: Miloš Zeman – 51.36%; Jiří Drahoš – 48.63%
- The second-round turnout: 66.6%
- Result: the incumbent president, Miloš Zeman, was re-elected

Candidates

Candidate	Features
Jiří Drahoš	An academic worker; a strongly apolitical candidate
Pavel Fischer	An ambassador to France and a former advisor of Václav Havel
Petr Hannig	A singer, producer, and talent-seeker
Marek Hilšer	A doctor and civil activist
Michal Horáček	A bookmaker, a lyric-writer, a Velvet Revolution symbol
Jiří Hynek	A businessman in weapon industry
Vratislav Kulháněk	A businessman connected with Škoda and Czech ice hockey
Mirek Topolánek	A former right-wing politician and manager
Miloš Zeman	A left-wing politician and economist



Investigation 1: New-Year Speeches

- 8 new-year speeches elaborated by the candidates on the request of Český rozhlas
- Miloš Zeman's authentic 2017 Christmas Address
- Reason of the choice: comparability of the samples
- Stylometric indexes
 - MATTR – Moving-Average Type-Token Ratio
 - ATL – Average Tokens Length
 - TC – Thematic Concentration
 - STC – Secondary Thematic Concentration
 - Q – Activity
 - VD – Verb Distances

Investigation 1: New-Year Speeches

Index	Formula	Interpretation
MATTR	$MATTR(L) = \frac{\sum_{i=1}^{N-L} V_i}{L * (N - L + 1)}$	Length-independent vocabulary richness of a text
ATL	$ATL = \frac{\sum L_{word}}{N}$	Lexis complexity
TC	$TC = \sum_{r=1}^T 2 * \frac{(h - r') * f(r')}{h * (h - 1) * f(1)}$	Thematic concentration of a text, and its thematic words
STC	$STC = \sum_{r=1}^T 2 * \frac{(2h - r') * f(r')}{h * (2h - 1) * f(1)}$	The same as TC, using 2h-point
Q	$Q = \frac{V}{A + V}$	POS-determined activity of a text
VD	$VD = \frac{\sum D_{verbs}}{N_{distance}}$	Syntactic complexity of a text

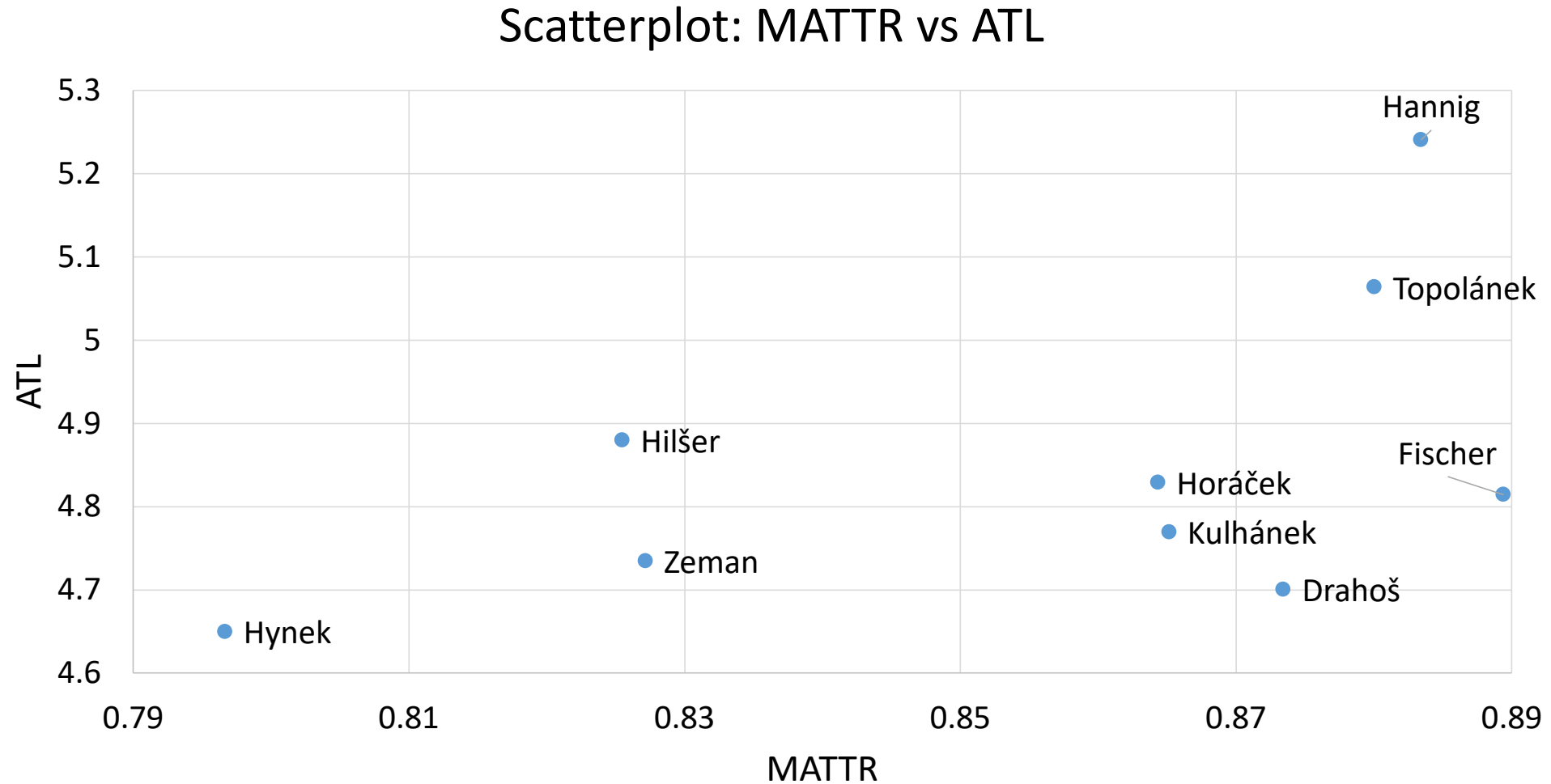
Investigation 1: New-Year Speeches

	MATTR	ATL	TC	STC	Q	VD
Drahoš	0.87	4.70	0.04	0.05	0.57	5.05
Fischer	0.89	4.81	0	0.09	0.46	6.36
Hannig	0.88	5.24	0	0.04	0.34	6.78
Hilšer	0.83	4.88	0	0.02	0.47	6.13
Horáček	0.86	4.83	0	0	0.56	4.81
Hynek	0.80	4.65	0	0.03	0.57	4.93
Kulhánek	0.87	4.77	0	0.03	0.58	4.23
Topolánek	0.88	5.06	0.09	0.11	0.46	5.42
Zeman	0.83	4.73	0	0.01	0.51	4.49

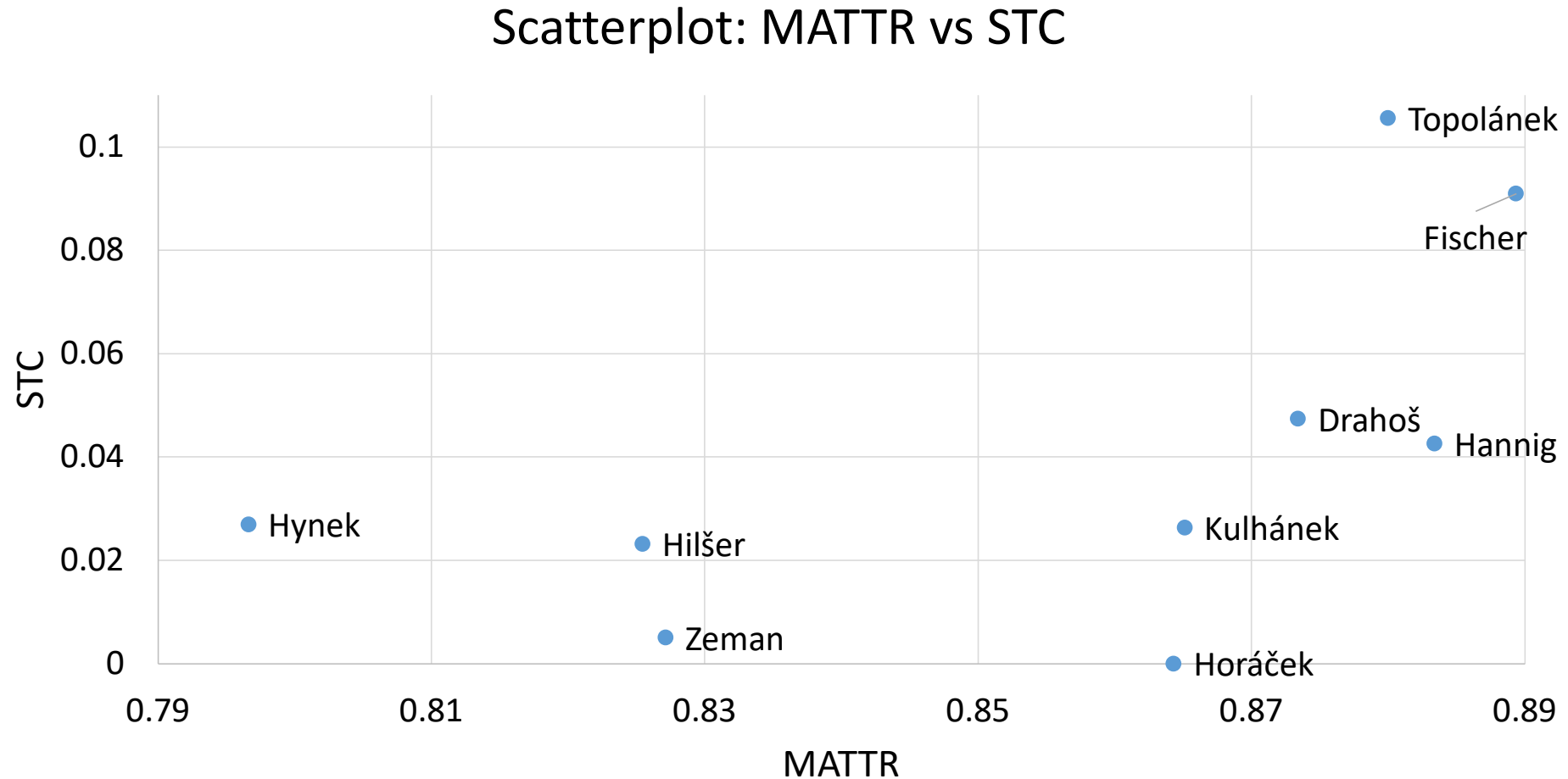
Investigation 1: New-Year Speeches

	MATTR	ATL	TC	STC	Q	VD
Drahoš	0.87	4.70	0.04	0.05	0.57	5.05
Fischer	0.89	4.81	0	0.09	0.46	6.36
Hannig	0.88	5.24	0	0.04	0.34	6.78
Hilšer	0.83	4.88	0	0.02	0.47	6.13
Horáček	0.86	4.83	0	0	0.56	4.81
Hynek	0.80	4.65	0	0.03	0.57	4.93
Kulhánek	0.87	4.77	0	0.03	0.58	4.23
Topolánek	0.88	5.06	0.09	0.11	0.46	5.42
Zeman	0.83	4.73	0	0.01	0.51	4.49

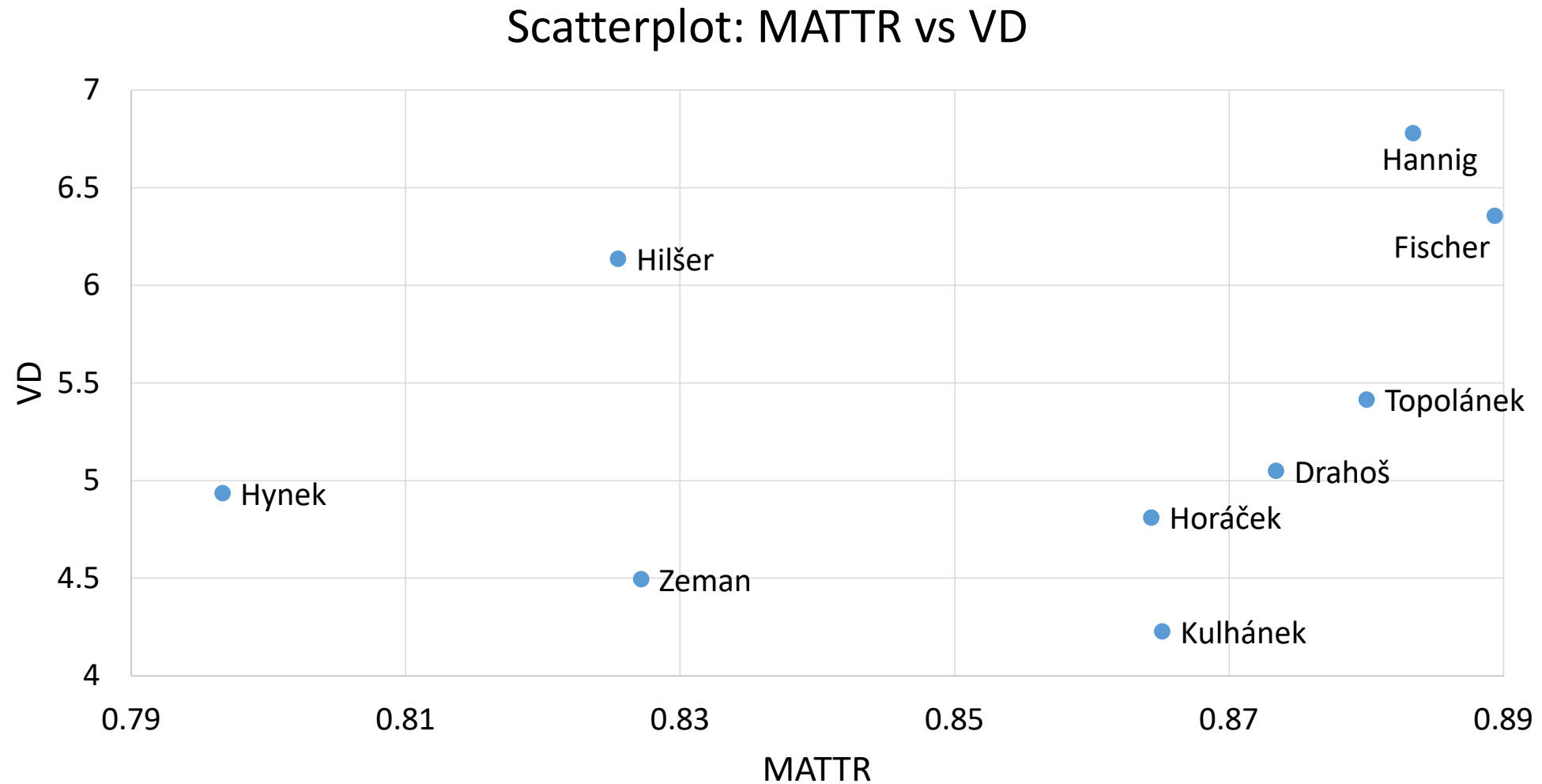
Investigation 1: Scatterplots



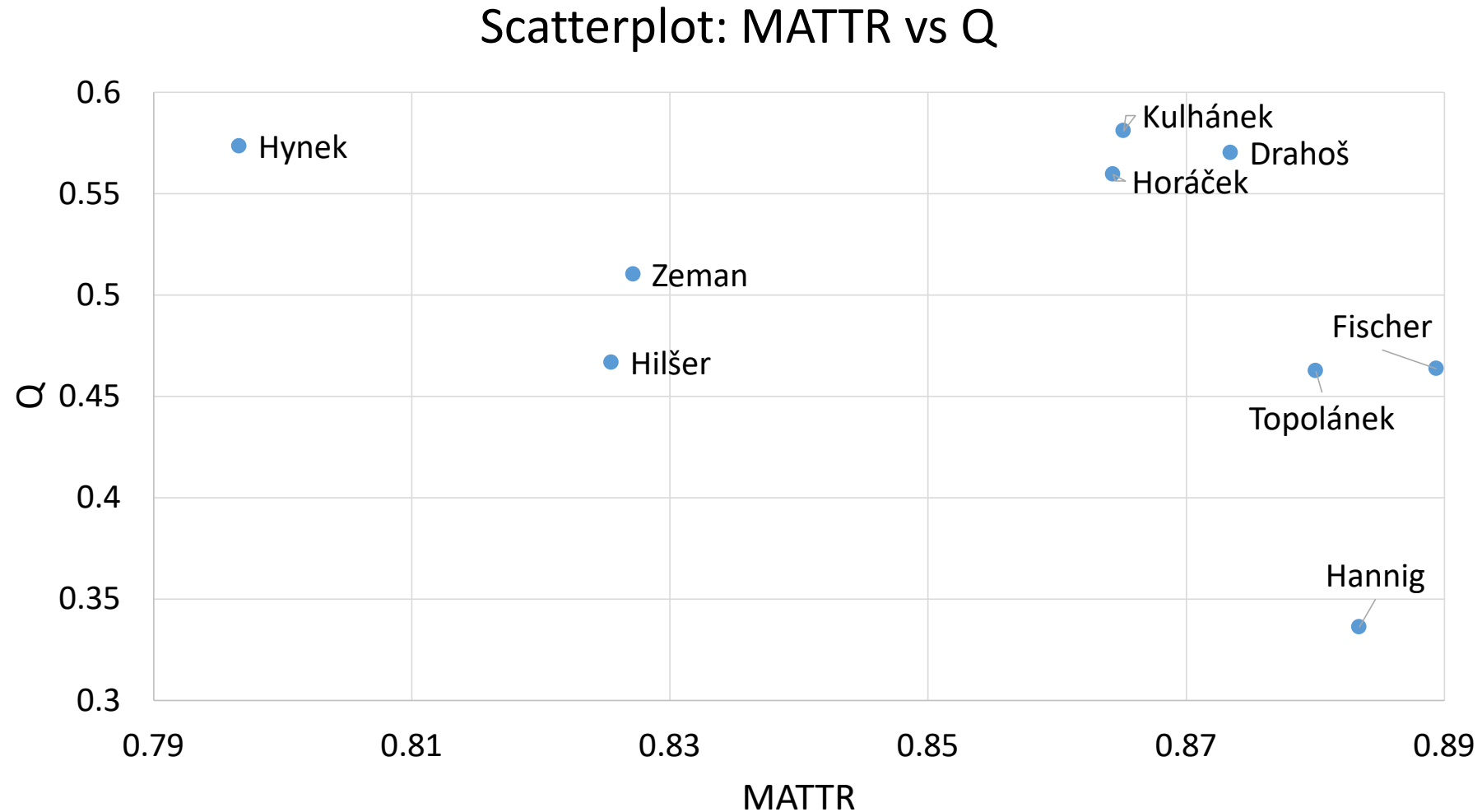
Investigation 1: Scatterplots



Investigation 1: Scatterplots



Investigation 1: Scatterplots



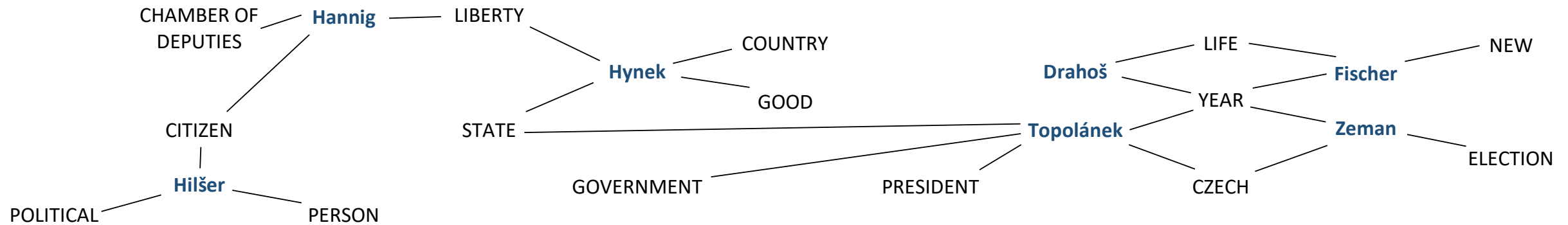
Investigation 1: Scatterplots

	MATTR	VD	ATL	Q	STC	Secondary Thematic Words	Outcomes
Drahoš	Higher	Lower	Lower	Higher	Middle	year; life	rich and active vocabulary; simpler style; theme-focus
Fischer	Higher	Higher	Lower	Middle	Higher	new; year; life	rich vocabulary; complex syntax; moderate activity; high theme-focus
Hannig	Higher	Higher	Higher	Lower	Middle	liberty; Chamber of Deputies; state; citizen	rich and descriptive vocabulary; complex style; theme-focus
Hilšer	Lower	Higher	Lower	Middle	Middle	political; citizen; person	simpler vocabulary and style; moderate activity and theme-focus
Horáček	Higher	Lower	Lower	Higher	Lower	–	rich and active vocabulary; simpler style; no theme-focus
Hynek	Lower	Lower	Lower	Higher	Middle	state; good; country	simpler, but active vocabulary and style; theme-focus
Kulhánek	Higher	Lower	Lower	Higher	Middle	country	rich and active vocabulary; simpler style; theme-focus
Topolánek	Higher	Lower	Higher	Middle	Higher	year; president; government; Czech; state	rich and complex vocabulary; simpler style; moderate activity; high theme-focus
Zeman	Lower	Lower	Lower	Middle	Lower	year; election; Czech	simpler vocabulary and style; moderate activity and lower theme-focus

Investigation 1: Scatterplots

	MATTR	VD	ATL	Q	STC	Secondary Thematic Words	Outcomes
Drahoš	Higher	Lower	Lower	Higher	Middle	year; life	rich and active vocabulary; simpler style; theme-focus
Fischer	Higher	Higher	Lower	Middle	Higher	new; year; life	rich vocabulary; complex syntax; moderate activity; high theme-focus
Hannig	Higher	Higher	Higher	Lower	Middle	liberty; Chamber of Deputies; state; citizen	rich and descriptive vocabulary; complex style; theme-focus
Hilšer	Lower	Higher	Lower	Middle	Middle	political; citizen; person	simpler vocabulary and style; moderate activity and theme-focus
Horáček	Higher	Lower	Lower	Higher	Lower	–	rich and active vocabulary; simpler style; no theme-focus
Hynek	Lower	Lower	Lower	Higher	Middle	state; good; country	simpler, but active vocabulary and style; theme-focus
Kulháněk	Higher	Lower	Lower	Higher	Middle	country	rich and active vocabulary; simpler style; theme-focus
Topolánek	Higher	Lower	Higher	Middle	Higher	year; president; government; Czech; state	rich and complex vocabulary; simpler style; moderate activity; high theme-focus
Zeman	Lower	Lower	Lower	Middle	Lower	year; election; Czech	simpler vocabulary and style; moderate activity and lower theme-focus

Secondary Thematic Word Network

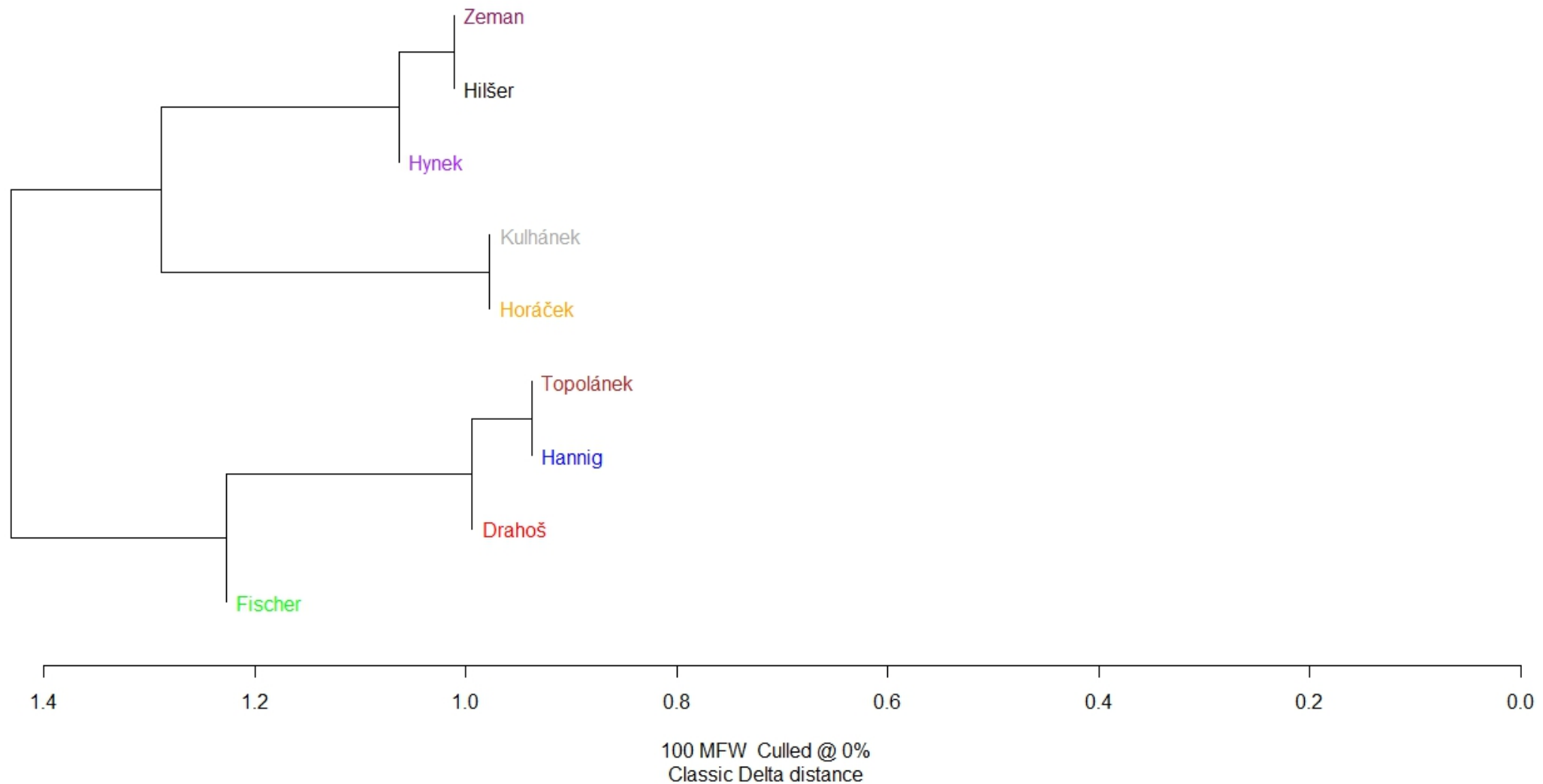


Investigation 2: MFW Analysis

- Most Frequent Words Analysis
- Based on comparing the given number of words in texts (here: 100)
- R Software; STYLO Package
- Visualisation: dendrogram
- Interpretation: sharing the topics among individual candidates

Investigation 2: Dendrogram

stylo
Cluster Analysis



Investigation 2: Interpretations

- **Zeman – Hilšer:** Focus on current issues (unemployment, economy, migration, citizenship, political situation)
- **Hynek:** close to Zeman – Hilšer; emphasis on the 1918 centenary
- **Kulhánek – Horáček:** Focus on the craftsmanship and nature of the Czech people
- **Topolánek – Hannig:** Focus on the election and political development
- **Fischer:** Confrontation with the incumbent president, generalities
- **Drahoš:** in between Fischer and Topolánek – Hannig (rhetoric questions with allusions + political development)

Investigation 3: Contrasting Programmes

- Two programmes: Marek Hilšer and Michal Horáček
- Reasons: Extensive enough
 - Hilšer: 2,612 tokens (target corpus)
 - Horáček: 6,969 tokens (reference corpus)
- Method: Keyword analysis; simple maths parameter (the ratio of relative frequencies)
- Products: positive keywords; lockwords; negative keywords

Investigation 3: Contrasting Programmes

413,48/396,04 per 10k-

Keywords +

péče
zdravotnictví
kvalitní
hrad
roli
výzkumu
najít
bude-li
ho
jmenování

Lockwords

byla
deset
ke
lépe
mimo
moci
moderní
nad
nato
našeho

Keywords -

jeho
let
nezbytné
méně
pak
právě
tu
ty
000
kdy

Hilser Heráček

Investigation 3: Contrasting Programmes

Positive Keywords	Lockwords	Negative Keywords
care	was [FEM]	his
healthcare	ten	years [GEN]
of-quality	to(wards)	vital [NEU]
castle	better [ADV]	less [ADV]
role [ACC; LOC]	except	then
research [LOC]	power [LOC]	just
find	modern	this [FEM]
if it is	over	these
him	afterwards	[numbers]
appointing	our [GEN]	when

Investigation 3: Interpretations

- **Hilšer**: focus on healthcare, research, and the role of the Castle
- **Horáček**: no particular focus in general; using numbers and functional words (confirms his tendency to no TC)
- **Lockwords**: attention paid to ourness and modernity

Investigation 4: Presidential Debate

- Material: The Czech TV Presidential Debate on Jan 25, 2018
- Corpora: Answers by Miloš Zeman and Jiří Drahoš
- Method: contrasting the corpora on the basis of keyword analysis
- Zeman: 3,157 tokens (target)
- Drahoš: 3,333 tokens (reference)



Investigation 4: Presidential Debate

275,67/317,94 per 10k-

Keywords +

právě
samozřejmě
řekl
tento
tedy
drahoši
dobře
mně
mě
otázku

Lockwords

allgemeine
anebo
bis
brýle
debaty
dnes
dotací
druhou
dále
eurem

Keywords -

øekl
miloš
zeman
takového
tøeba
pøípadì
samozøejmì
urèitì
øíkám
dobøe



Investigation 4: Presidential Debate

Positive Keywords	Lockwords	Negative Keywords
just	allgemeine	Miloš
of course	or	Zeman
[he] said	BIS	such a thing
this one [MASC]	glasses	perhaps
so	debate [GEN]	case [LOC]
Drahoš [VOC]	today	surely
well	donation [GEN PL]	I say
me [DAT; LOC]	second [INST]	
me [GEN; ACC]	further	
answer [ACC]	Euro [INST]	

Investigation 4: Interpretations

- Both candidates react to each other, but in a different way
 - Zeman: direct address (vocative case)
 - Drahoš: reference (nominative case)
- Speaking about themselves, they use distinct strategies
 - Zeman: indirect reference (*me*)
 - Drahoš: direct reference (*I*)
- Drahoš: a prominent use of *perhaps*
- Lockwords
 - Confrontations (*glasses*)
 - Common topics (*Euro*, *BIS* – Czech Intelligence Agency, *allgemeine* – reference to an article in a German newspaper)
 - Focusing on the current issues (*today*, *debate*)

Investigation 4: Miscellaneous

Text	TTR	ATL	TC	STC	Q	VD
Drahoš	0.278	4.720	0.008	0.018	0.621	4.322
Zeman	0.309	4.814	0.005	0.007	0.592	4.789

- QUITA indexes
- Considerable differences in VD
- Drahoš: more concentrated as to STC

Investigation 4: Thematic and Secondary Thematic Words

<i>Drahoš</i>		<i>Zeman</i>	
TW	STW	TW	STW
president	president	Mr	Mr
say	say		say
Mr	Mr		Mrs
	say [impf]		year
	Miloš		go
	thing		
	person		
	government		

- Confirms the reaction-like character of the debate (*Mr* + names; *say*)
- Drahoš: using a lot of vague thematic words (*thing*, *person*)
- Zeman: focusing on facts (*year*) and active voting participation (*go*)

Summary

Candidate	Style Feature	Topics
<i>Drahoš</i>	Dynamics	Politics
<i>Fischer</i>	Intellectualism; focus	Politics; confrontation
<i>Hannig</i>	Art; congestion	Politics
<i>Hilšer</i>	Moderation; simplicity	Economy and society
<i>Horáček</i>	Dynamics; lack of focus	Nation
<i>Hynek</i>	Simplicity	Economy and society; 1918 anniversary
<i>Kulhánek</i>	Dynamics	Nation
<i>Topolánek</i>	Intellectualism; focus	Politics
<i>Zeman</i>	Simplicity	Economy and society

- Opening: analysis of newspaper interviews + more material in general