

# Contrasting (mainly) Coreference In Czech And German

**From Different Frameworks to Joint Results**

GECCO-PDĪT

or

PDT-GECCO comparison during the last two years

*Anna Nedoluzhko, Katja Lapshinova-Koltunski*

# What we have:

- two large-scale annotated corpora



GECCO



PDT 3.0

## Motivation - general

- compare languages (to know more about each)
- area of discourse phenomena
  - connectives
  - coreference → coreference chains
- variation of coreference chains reflects cohesion

## Motivation - material

- take advantage of the existing resources reflecting systemic peculiarities of Czech and German
- use Czech and German data annotated on the basis of two different frameworks



*General comparison*

comparable and parallel texts in English and German (ca. 80,000 sentences)



texts in Czech (ca. 50,000 sentences)

written and spoken texts



written texts

different genres and registers



journalistic texts with further genre classification

morpho-syntactic annotation



multilayer annotation:  
morphological, analytical and  
tectogrammatical



*Discourse phenomena comparison*

cohesive devices



textual phenomena

(co)reference

coreference

lexical cohesion

bridging

substitution

-

ellipsis

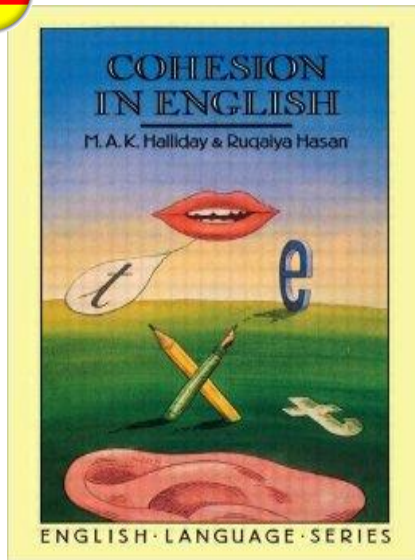
ellipsis captured in dependency trees

conjunctives

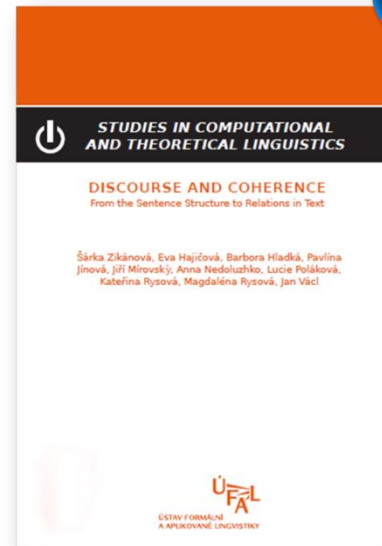
connectives, arguments, relations



*two different theoretical approaches*

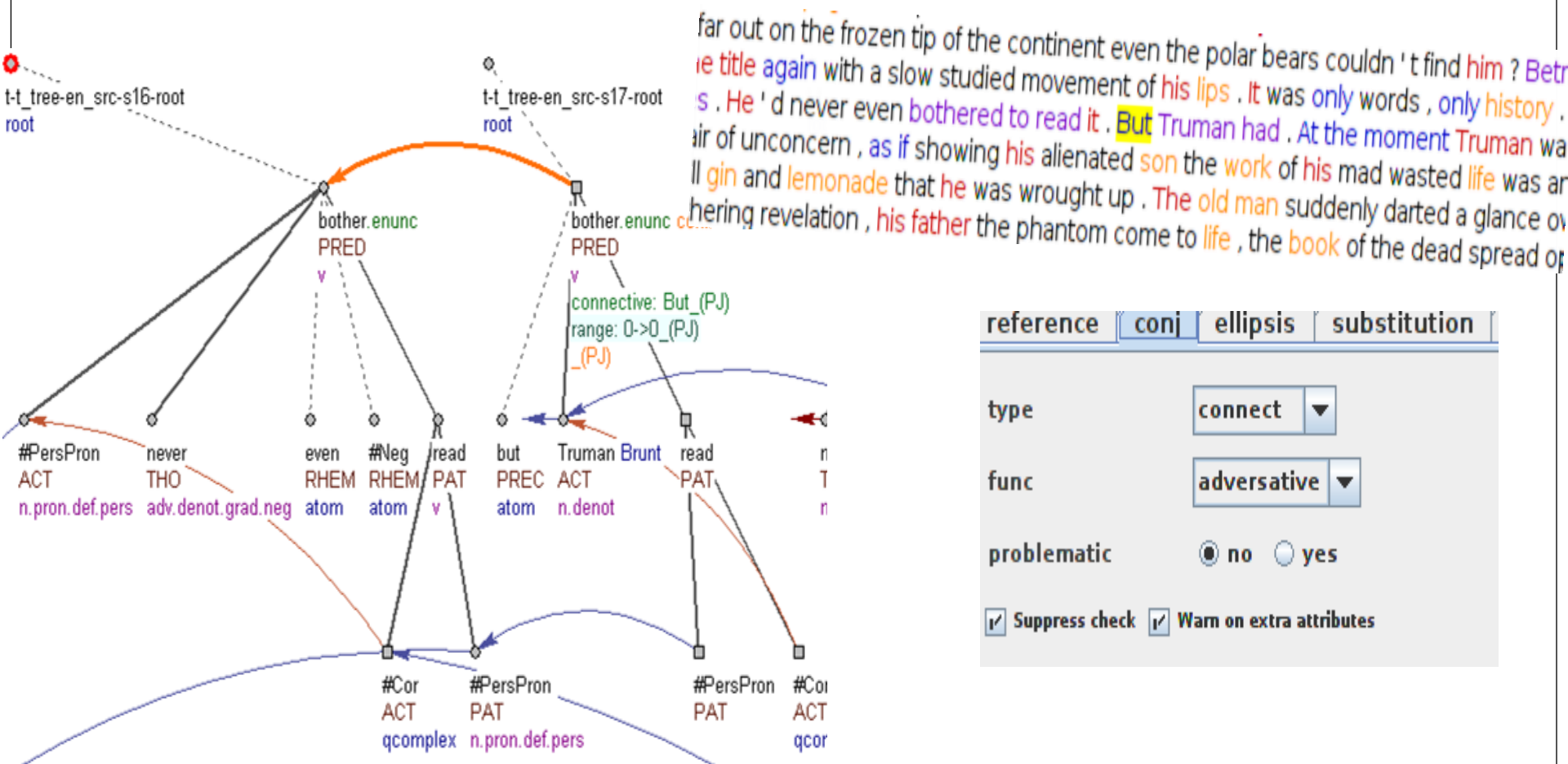


Halliday and Hasan, 1976



Sgall 1986, Functional grammar,  
Zikánová et al., 2015  
PDTB-like discourse

# Visualisation in Tools



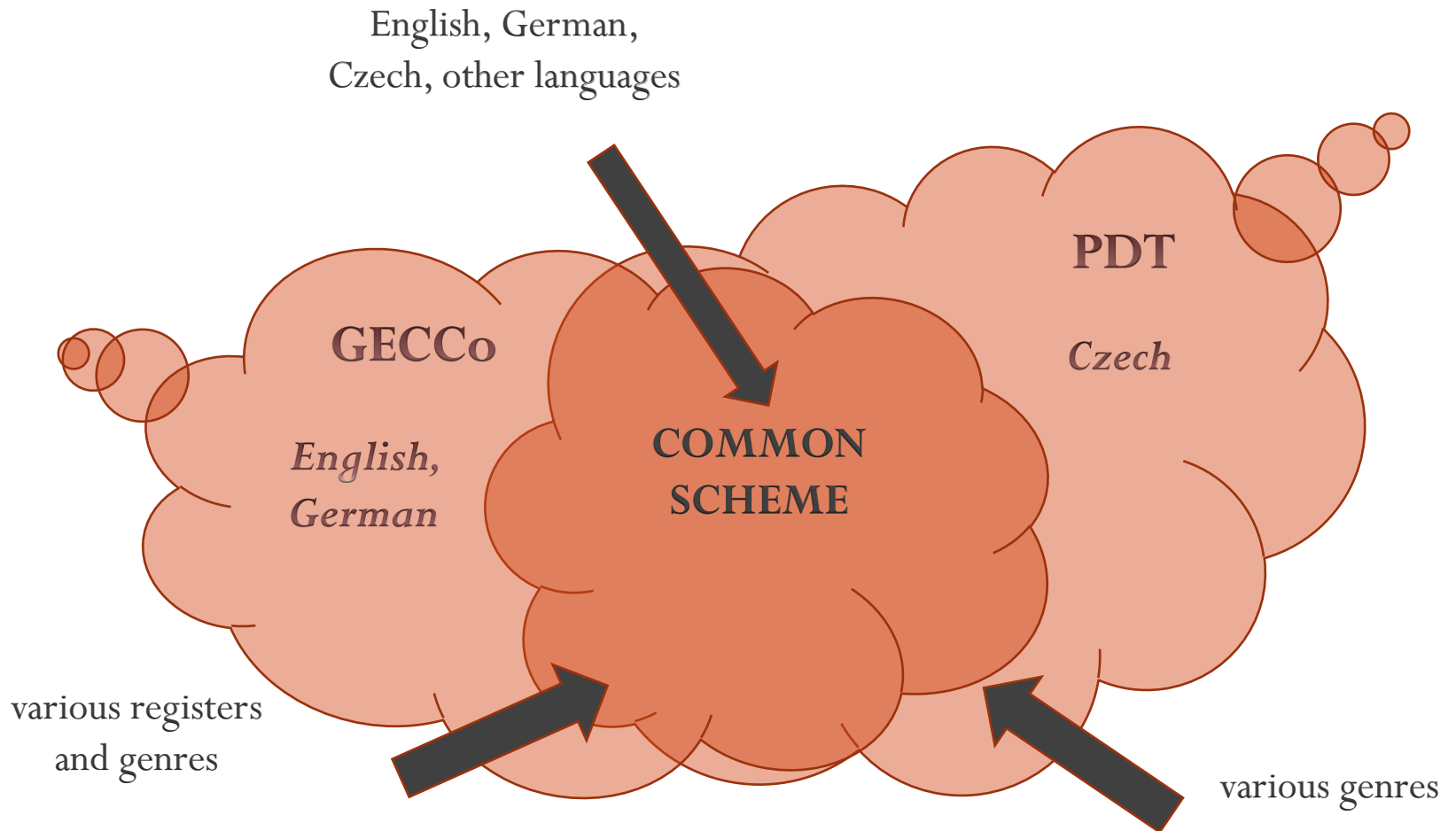
far out on the frozen tip of the continent even the polar bears couldn 't find him ? Betr  
 ie title again with a slow studied movement of his lips . It was only words , only history .  
 s . He 'd never even bothered to read it . But Truman had . At the moment Truman wa  
 air of unconcern , as if showing his alienated son the work of his mad wasted life was ar  
 ll gin and lemonade that he was wrought up . The old man suddenly darted a glance ov  
 hering revelation , his father the phantom come to life , the book of the dead spread of

reference	conj	ellipsis	substitution
type	connect ▼		
func	adversative ▼		
problematic	<input checked="" type="radio"/> no <input type="radio"/> yes		
<input checked="" type="checkbox"/> Suppress check	<input checked="" type="checkbox"/> Warn on extra attributes		

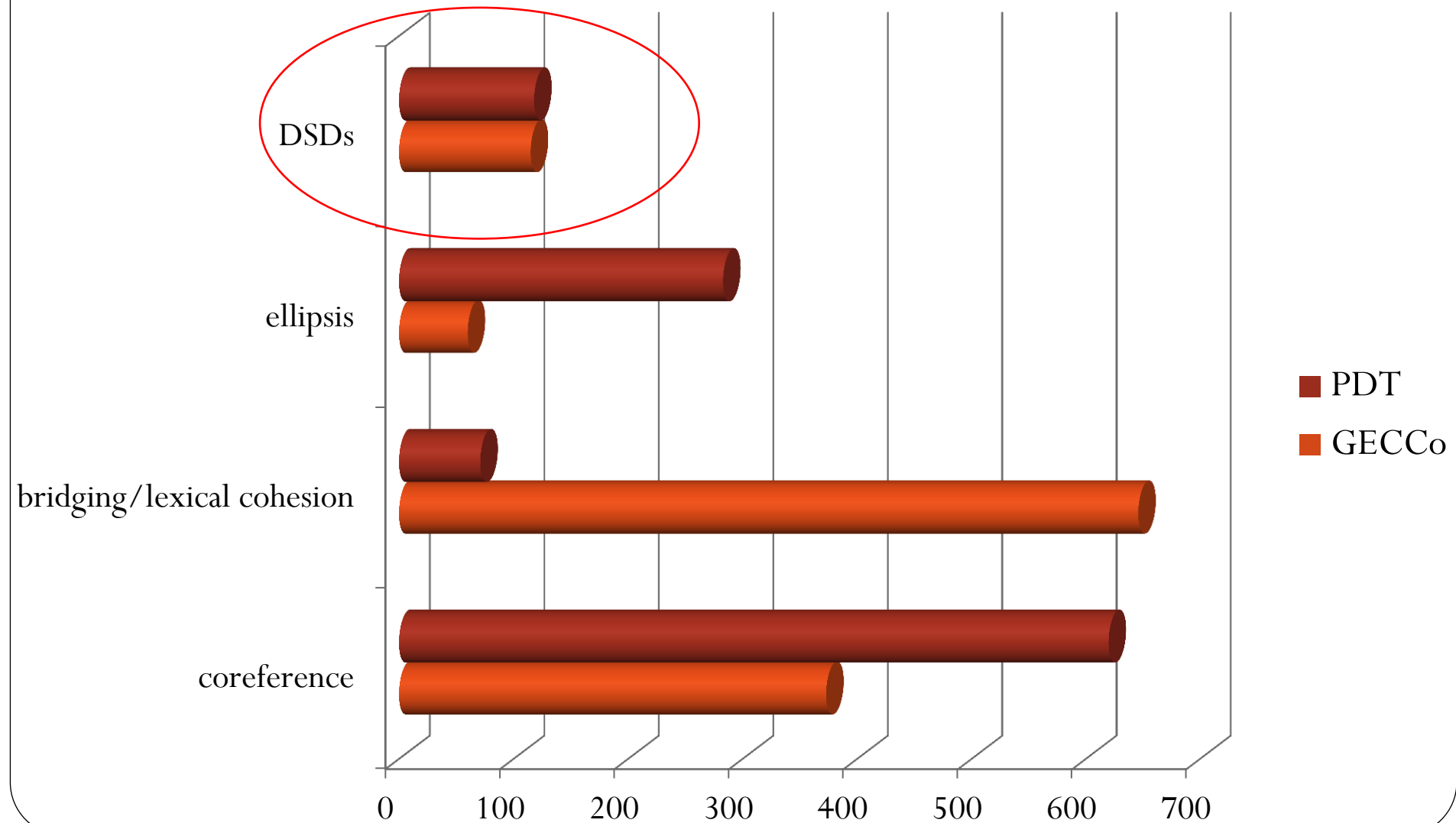
He 'd never even bothered to read it. **But** Truman had.



# Creating a common scheme



# Discourse Structuring Devices

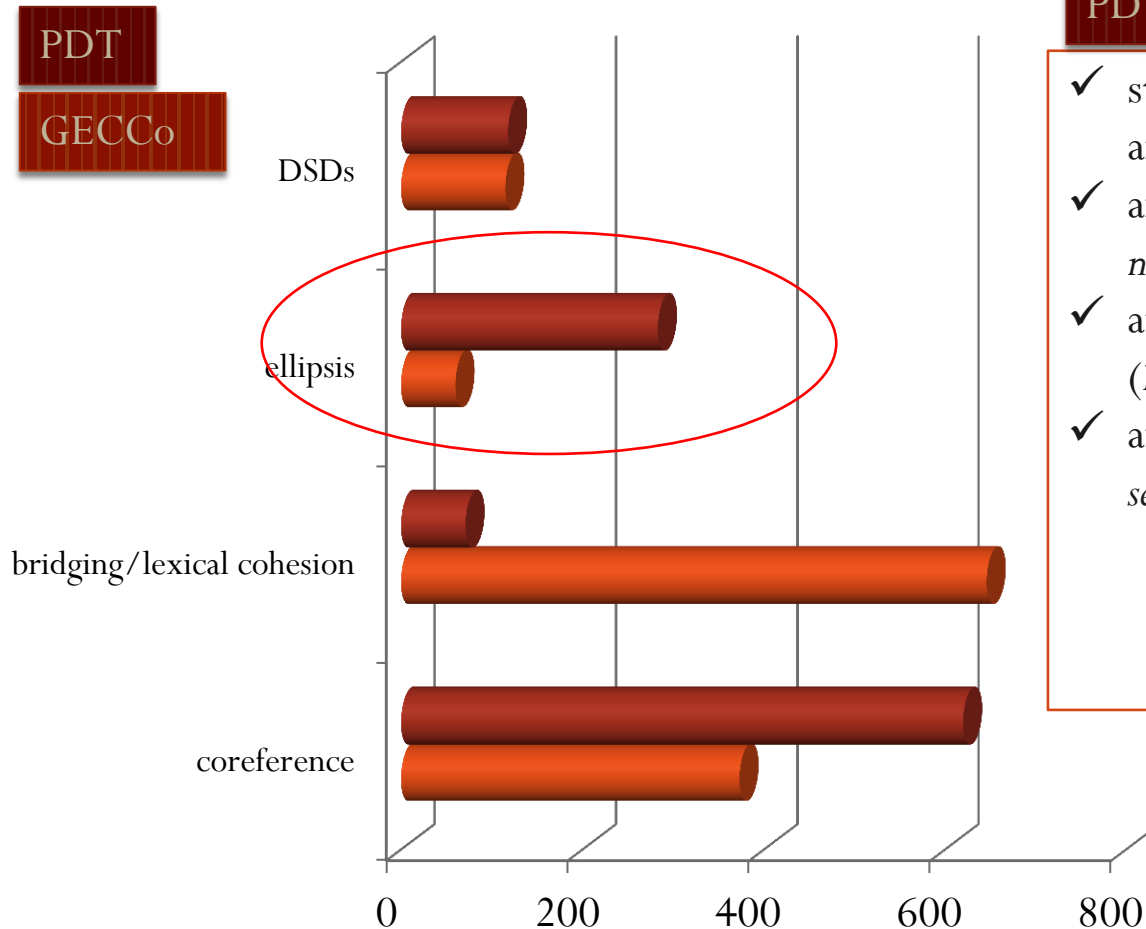


Preced. –succ.	reason - result	confrontation	conjunction
synchronous	Pragm. reas. – result	opposition	instantiation
	purpose	pragmatic contrast	specification
	explication	Restr. opp. + except.	equivalence
	condition	concession	generalization
	Pragm. cond.	correction (replacement)	conjunctive alternative
		gradation	disjunctive alternative
<b>TEMPORAL</b>	<b>CONTINGENCY</b>	<b>COMPARISON (CONTRAST)</b>	<b>EXPANSION</b>

*discourse markers  
(attitude markers, modal  
particles) -  
in PDT not considered as  
connectives*

<b>TEMPORAL</b>	<b>CAUSAL</b>	<b>ADVERSATIVE</b>	<b>ADDITIVE</b>	<b>MODAL</b>
temporal relation between events	relation of causality/dependence between	relation of contrast/alternative, for two events which are not true at the same time	relation of addition, for two events that are true/not true at the same time	relation between events connected by an evaluation of the speaker
<i>after, afterwards, at the same time..</i>	<i>because, therefore, that's why...</i>	<i>yet, although, by contrast...</i>	<i>and, in addition...</i>	<i>well, sure, of course, surely, eventually...</i>
<i>nachdem, danach, gleichzeitig..</i>	<i>weil, deshalb, aus diesem Grund..</i>	<i>doch, obwohl, im Gegensatz dazu..</i>	<i>und, außerdem..</i>	<i>klar, sicher, allerdings, jedenfalls, eigentlich, wohl...</i>

# Ellipsis



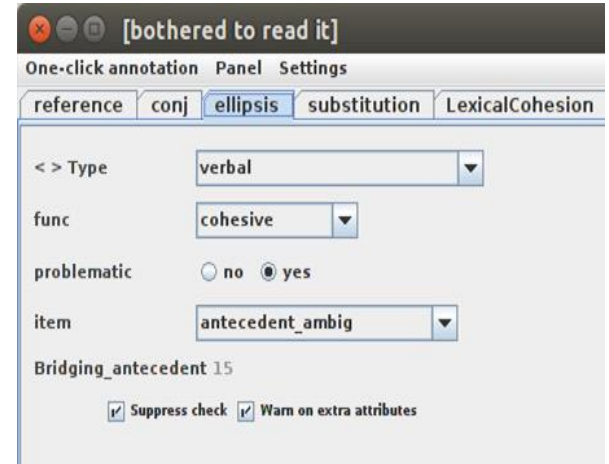
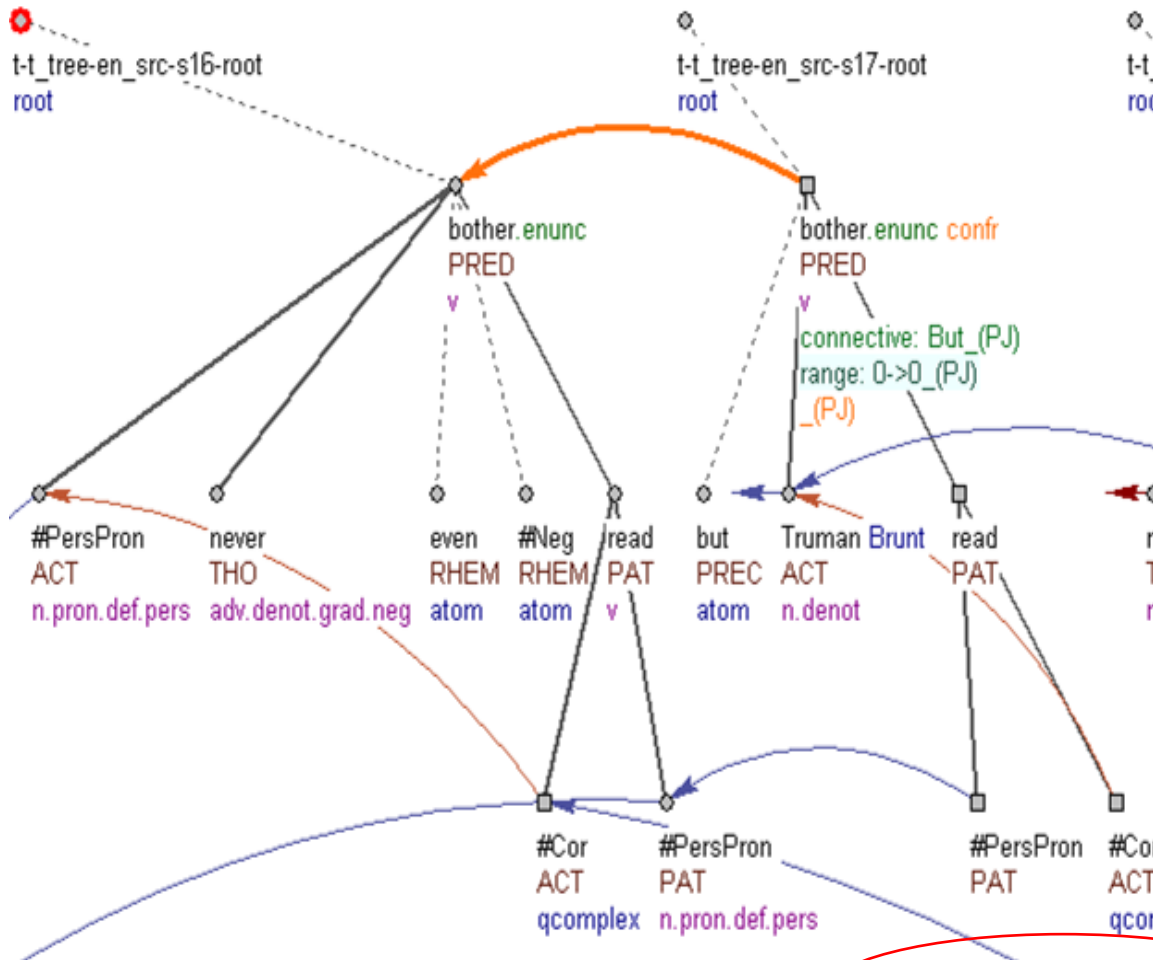
PDT

- ✓ structural ellipses (of governing verbs and nouns),
- ✓ anaphoric zeros (*Their reaction was 0 to do nothing and 0 ride it out*),
- ✓ arguments with control constructions (*Peter wants to [Peter] sleep*),
- ✓ arguments with general arguments (*Jane sells at Bata [what] [to whom]*)

GECCo

- ✓ only in the case of textual relations across grammatical domains

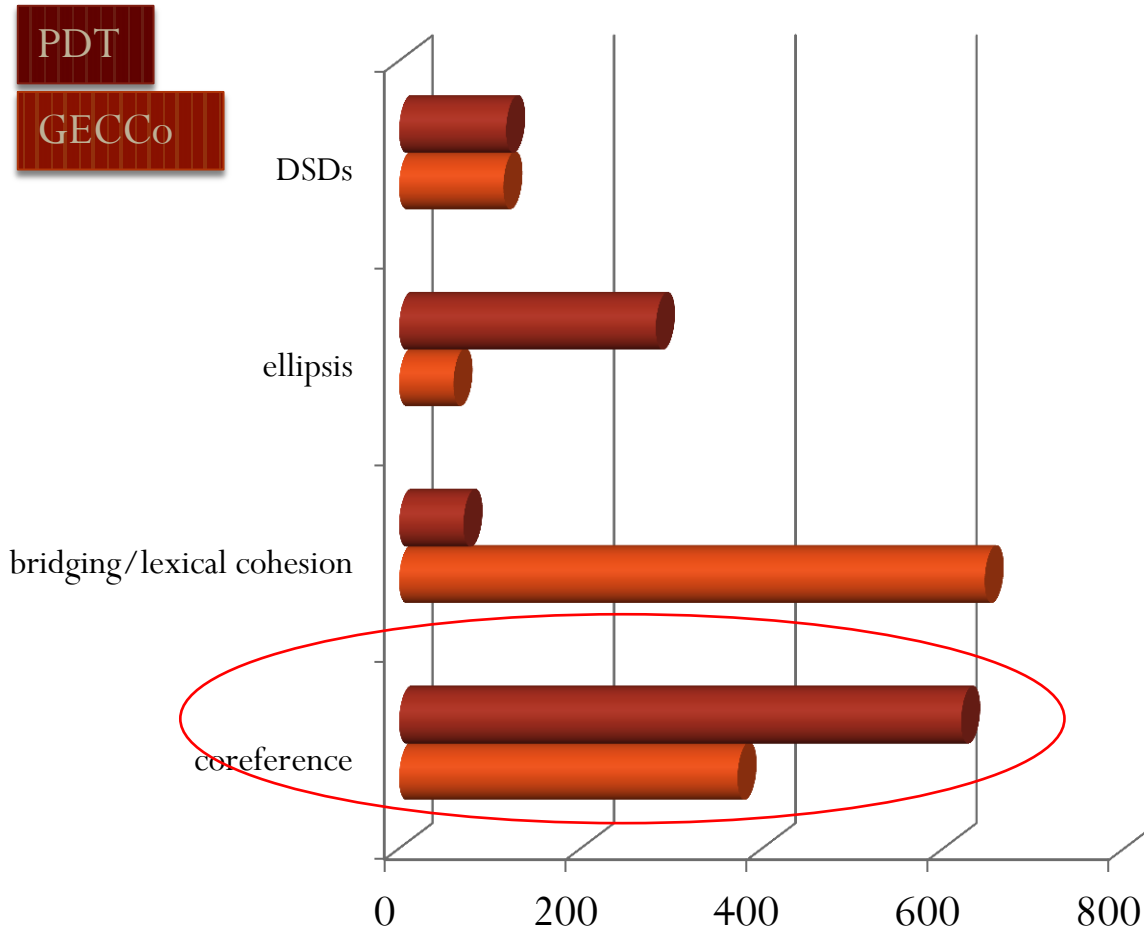
# Ellipsis in Both Approaches



out on the frozen up of the continent, even the polar bear  
 title again with a slow studied movement of his lips. It was  
 He'd never even bothered to read it. But Truman had  
 of unconcern, as if showing his alienated son the work of

*He'd never even bothered to read it. But Truman had [bothered to read it].*

# Coreference



PDT

- ✓ both implicit and explicit relations of coreference are annotated, including indefinite NPs

GECCo

- ✓ only with explicit marker

*John left Mary. He wanted to see his mother.*

*My daughter likes to dress herself without my help.*

*It means that he doesn't really love Mary.*

*Alex is the boy who kissed Mary.*

*If you want to see him, PRO take a look at his outfit.*

*At last, Mary came to Prague herself and found the Prague atmosphere quite casual.*

**PDT**

**GECCo**

**Coreference**

*John asked his mother to advise him how he should behave with Mary, but mother ignored his wish.*

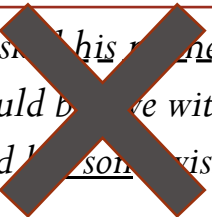
*John and Mary kissed Ø.PAT*

*Peter wanted Ø.ACT sleep.*

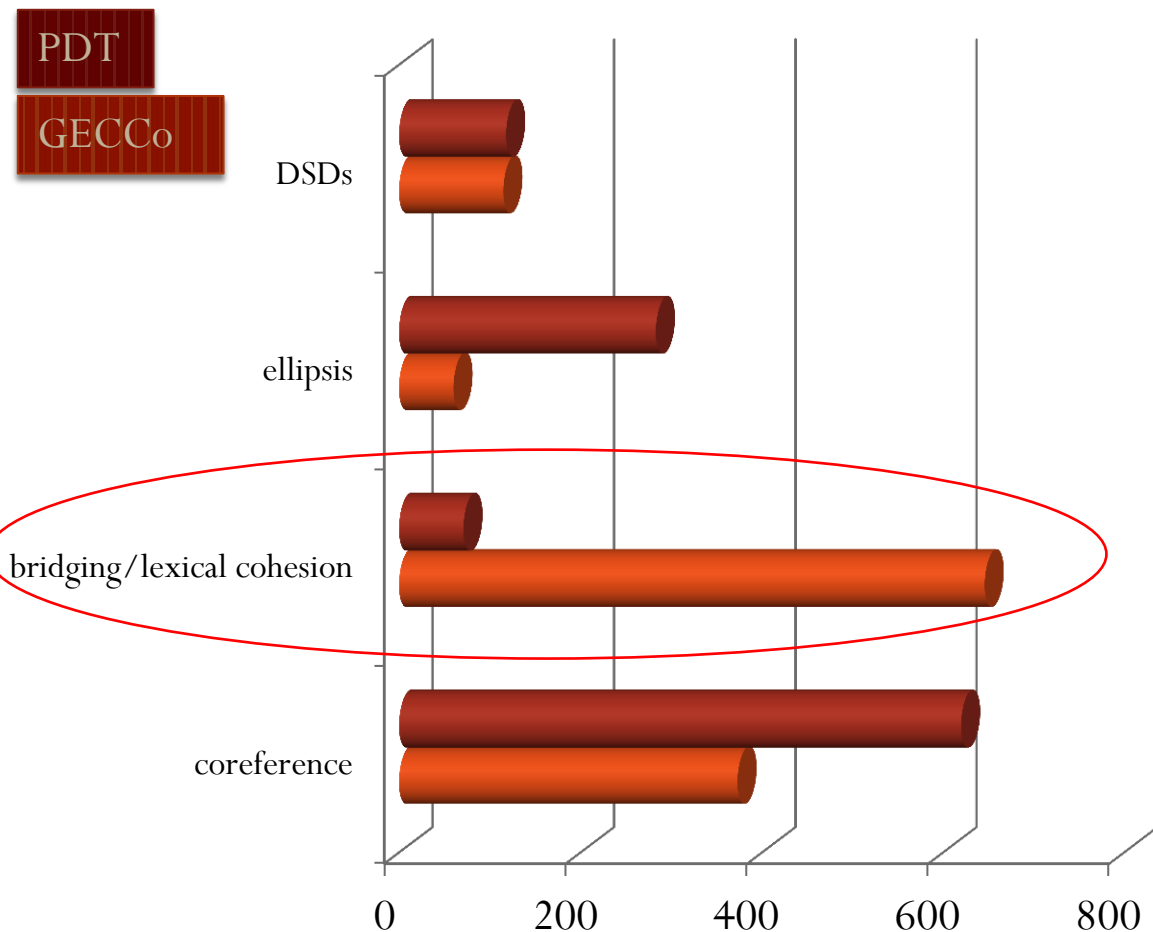
*John asked mother to come to Mary's place with him but she decided not to go there.*

*Mary suggested John to go to the theater, but John ignored her wish.*

*John and Mary Ø.ACT run around the lake.*



# Bridging Relations / Lexical Cohesion



PDT

- ✓ 6 specific types of semantic/pragmatic relations

GECCo

- ✓ many types with explicit marker



**repetition:** same word is repeated (*artist - Artist - artists*)

**synonym:** same meaning (*film - movie*)

**antonym:** opposite meaning (*day - night, husband - lover*)

**hypernym:** more generic word (*street people - vagabond*)

**GECCo**

## Bridging/Lexical Cohesion

**PDT**

**co-hyponym:** word with same degree of specificity (*movie director - producer - writer*)

**meronym:** a word that names a part of a larger whole (*film - scenes*)

**co-meronym:** a word that names another part of a larger whole (*scenes - characters*)

**holonym:** the larger whole (*film as related to scenes and characters*)

**PART-WHOLE** (*film - scenes*)

**SET – SUBSET/element of the set** (*the beds - one of the beds*)

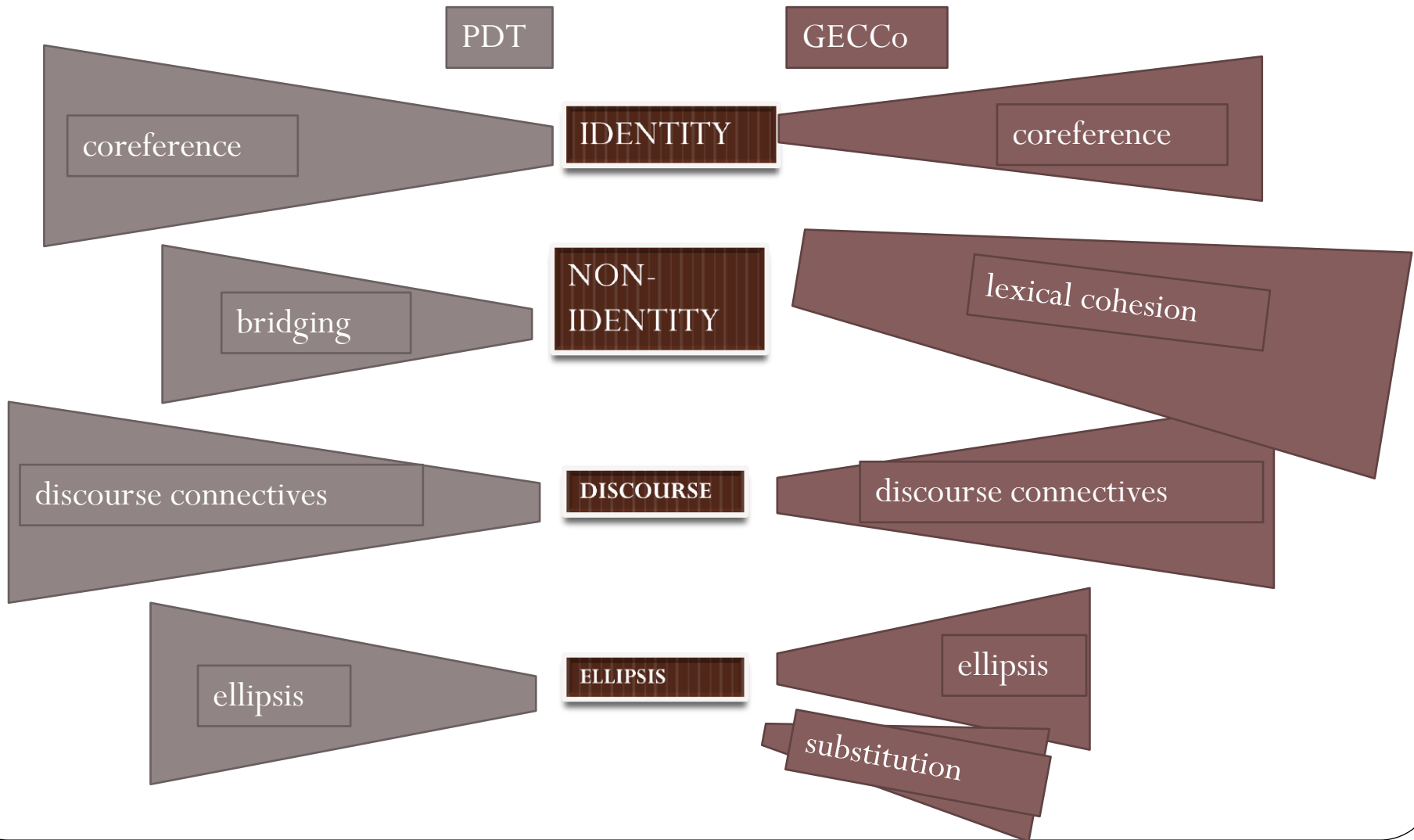
**function – object** (*prime-minister – government, trainer – football team*)

**semantic/contextual CONTRAST** (*It is necessary to have lofty aims and not to be satisfied with small (ones)*)

**explicit anaphora without coreference** (*world war – at that time*)

**REST** (family (*grandfather - grandson*), place – inhabitant, author – work, same denomination to support cohesion of the text (*a chance helped – another chance entered the game ...*) and event – participant of the event (*enterprise - entrepreneur*))

# MUTUAL RESULTING SCHEMA



# MUTUAL RESULTING SCHEMA

	featID	Czech	German
<b>IDENTITY</b>	id1	coreference with pronouns	coreference with heads (no extended reference)
	id2	pronouns with arrows to segments and events	reference to verb phrases and longer segments
	id3	NP coreference	coreference with modifiers or def.articles
	id4	coreference with the word <i>same</i>	general comp.reference
	id5	coreference with local and temporal adverbs	coreference with local and temporal adverbs
<b>NON-IDENTITY</b>	nonid1	relations of MERONYMY	relations of MERONYMY
	nonid2	bridging CONTRAST	particular comparative reference and antonyms
<b>DISCOURSE RELATIONS</b>	temp	temporal	temporal
	cont	contingency	causal
	comp	comparison (contrast)	adversative
	expan	expansion	additive
<b>ELLIPSIS</b>	ellipsis	textual ellipsis	cohesive ellipsis

# comp with discourse relations of comparison/contrast

**German:** *Arbeiten wie die Polen, aber leben wie die Japaner... [Work like the Poles, but live like the Japanese...]*

**Czech:** *Poslední statistické sčítání dopravy proběhlo v roce 1990. Za poslední tři roky se však na českých silnicích zvýšil provoz. [The latest statistical traffic census took place in 1990. Over the past three years, however, traffic on Czech roads has increased.]*

BUT: could be different

*William Gates and Paul Allen in 1975 developed an early language-housekeeper system for PCs, and Gates became an industry billionaire six years after IBM adapted one of these versions in 1981.*

(from experimental English excerption)

PDT: *reason-result*

GECCo: *additive relation*

# nonid1 with meronymy relations

**German:** ...*praktisch wird es dazu nicht kommen – dafür ist in Deutschland die Bereitschaft zur Solidarität, der Glaube an das “für alle” zu groß. Eine andere Gefahr ist da weit realer: daß die Deutschen...* [In practice... it will not come to that– the readiness to practice solidarity, and people’s belief in the “for all” is too pronounced in Germany. Another dangers much more real, however: that the Germans....]

**Czech:** *Jednotlivá studia v apartmánech jsou vybavena kuchyní, takže je možná individuální příprava stravy. [Studio apartments are equipped with kitchens, so everyone can prepare their own food.]*

# nonid2 with contrast relations

**German:** *Dazu gehören zum Beispiel die Halbierung der Energie- und Rohstoffintensität bis 2020 gegenüber 1990 und die Verdoppelung des Anteils erneuerbarer Energien am Energieverbrauch bis 2010. [For example, halving the amount of power and raw material consumption by 2020 compared to 1990 levels and doubling the percentage of renewable energy used as part of total energy consumption by 2010.]*

**Czech:** *Saldo běžného účtu platební bilance podle odhadu dosáhlo vloni cca 600 USD... I když letos a příští rok je nutné počítat se zpomalením růstu vývozu, prognózujeme, že saldo přesto zůstane kladné. [The balance of the current account deficit is estimated to reach \$600 last year... Although this and the next years we expect the slowdown in export growth, we forecast that the deficit will still remain positive.]*

# Nominal Ellipsis

**German:** *All das ist eine kleine Revolution. Die grössere [] ist diese: ... [But there is also a bigger [revolution], and it is this]*

**Czech:** *Klienti pojišťoven, které ukončí svou činnost, se automaticky vrátí k Všeobecné []. [Clients of insurance companies which shut down will automatically return to the General [one].]*



# Applications of the schema

- To calculate all kinds of textual/cohesive phenomena on an excerption of texts in Czech and German (Katja-Anja-Kerstin, Budapest&LREC 2016)
- To consider coreference chains in Czech and German (Anja&Katja, Dialog 2016)
- To consider abstract anaphora in Czech and German (Anja&Katja, CORBON, San Diego 2016)



GERMAN



CZECH

- reduced morphology
- more isolating
- less flexible word order
- definite article

- morphologically rich
- less isolating
- flexible word order
- no definite article

# Data Excerption

- written discourse
- essays
- comparable topics (economic, political and social issues)

# Data Excerption - 1

*For general textual phenomena (Budapest, LREC)*



id	topics	sent	tokens
1	Germany and social market economy	121	2035
2	Optimistic remarks on globalisation	47	971
3	Politics and globalisation	103	1871
4	Globalisation and new challenges	27	478
5	The biggest currency changeover	85	1460
6	Globalisation and market economy	80	1782
7	Global market and technical progress	108	1851
8	Economic and technological changes	73	1795
9	Doctors and medical system	92	2687
	<b>Total: ALL GERMAN</b>	<b>736</b>	<b>14930</b>



id	topics	sent	tokens
1-5	Germany, politics and history	170	687
6	Housing	83	1644
7-8	Technological changes	73	1795
9-12	Politics	121	1854
13-14	Economics	149	2568
15-16	Unemployment	112	2252
17	Television	55	969
	<b>Total: ALL CZECH</b>	<b>763</b>	<b>11769</b>

# Data Excerption - 2



	<b>Topics</b>	<b>sent</b>	<b>tokens</b>
1	Germany and social market economy	121	2035
2	Optimistic remarks on globalisation	47	971
3	Politics and globalisation	103	1871
4	Globalisation and new challenges	27	478
5	The biggest currency changeover	85	1460
6	Globalisation and market economy	80	1782
7	Global market and technical progress	108	1851
8	Economic and technological changes	73	1795
	<b>Total: ALL GERMAN</b>	<b>645</b>	<b>12243</b>

*For coreference (Moscow, USA)*

written essays with comparable topics on economic, political and social issues



<b>id</b>	<b>topics</b>	<b>sent</b>	<b>tokens</b>
1-5	Germany, politics and history	170	687
6	Housing	83	1644
7	Technological changes	73	1795
8-9	Politics	121	1854
10-13	Economics	149	2568
14	Unemployment	112	2252
15	Television	55	969
	<b>Total: ALL CZECH</b>	<b>628</b>	<b>11399</b>

# Analysis on Data Excerption – 1

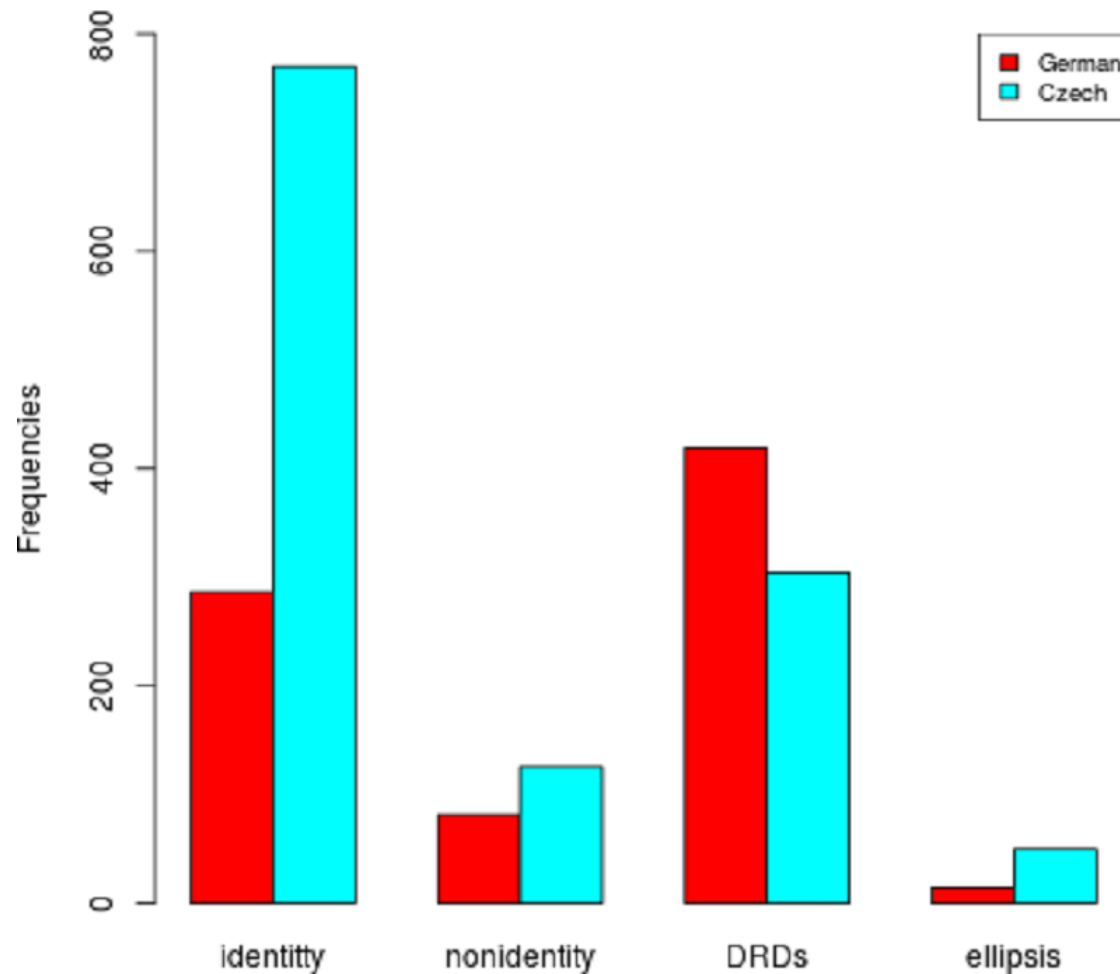
## Subcategories (*norm. per 1000*)

featID	id1	id2	id3	id4	id5	nonid1	nonid2	temp	cont	comp	expan	ellipsis
German	88.4	38.2	144.7	3.3	12.0	52.9	28.8	106.5	52.2	79.0	181.5	14.1
Czech	97.7	64.6	597.3	0.0	10.2	88.4	37.4	14.4	66.3	86.7	136.8	50.1

	featID	Czech	German
<b>IDENTITY</b>	id1	coreference with pronouns	coreference with heads (no extended reference)
	id2	pronouns with arrows to segments and events	reference to verb phrases and longer segments
	id3	NP coreference	coreference with modifiers or def.articles
	id4	coreference with the word <i>same</i>	general comp.reference
	id5	coreference with local and temporal adverbs	coreference with local and temporal adverbs
<b>NON-IDENTITY</b>	nonid1	relations of MERONYMY	relations of MERONYMY
	nonid2	bridging CONTRAST	particular comparative reference and antonyms
<b>DISCOURSE RELATIONS</b>	temp	temporal	temporal
	cont	contingency	causal
	comp	comparison (contrast)	adversative
	expan	expansion	additive
<b>ELLIPSIS</b>	ellipsis	textual ellipsis	cohesive ellipsis

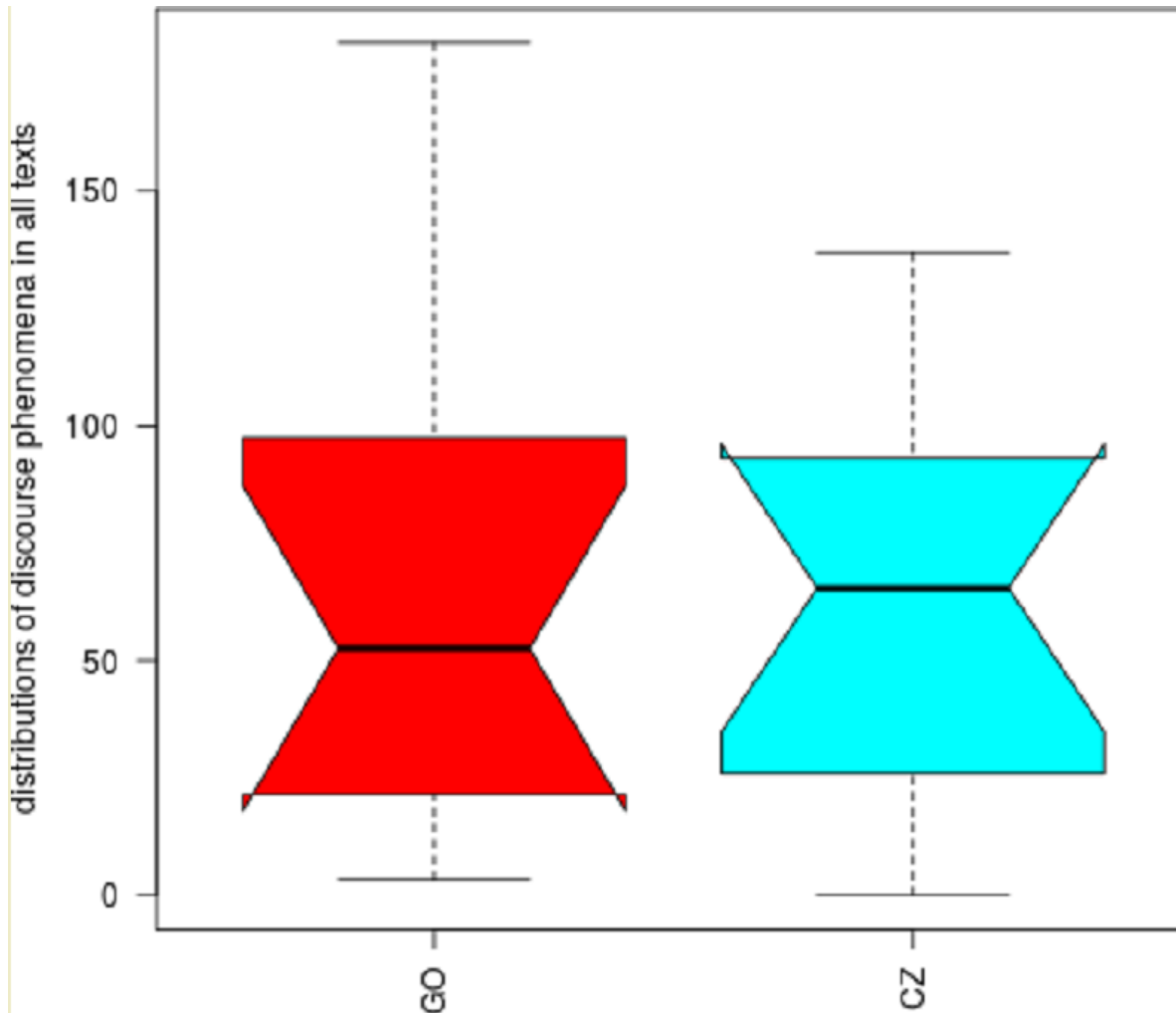
# Textual phenomena on Data

## Excerption - 1



# Textual phenomena on Data

## Excerption - 1

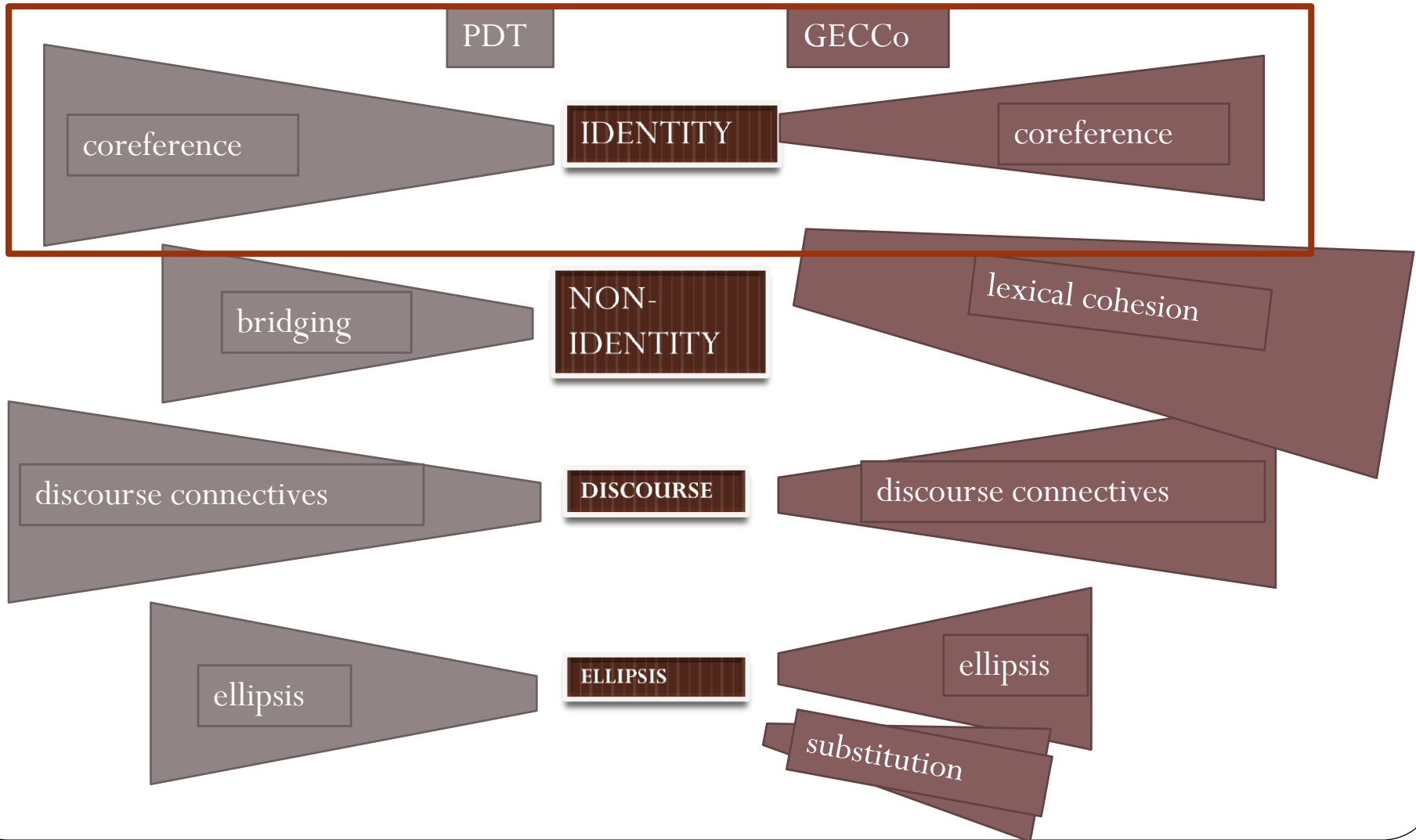




# RESULTS

- interoperable schemes permit multilingual analyses of resources from different approaches, saves time and effort as no additional annotation is required
- insights into differences between German and Czech
- limitations of the dataset: although comparable, still different sources, authors, size
- numbers show more differences in conceptualizations and annotation details, not in languages so far

# MUTUAL RESULTING SCHEMA



# Coreference chains and abstract anaphora

- Are there any language contrasts with respect to coreference chain features between German and Czech?
- Are there any differences between German and Czech if abstract coreference is concerned?

# Features to answer:

- longest chains (chains with the greatest number of elements in a text),
- number of coreference pairs (= number of anaphorically referring expressions),
- types of anaphors (e.g. pronouns, nominal modifiers, temporal and local adverbs and so on),
- number of anaphora referring to antecedents other than NP/pronoun,
- types of anaphora referring to antecedents other than NP/pronoun

# Chain properties in German and Czech

	GERMAN	CZECH
number of coreference chains	225	550
length of coreference chains (mean value per text)	2.45	3.2
longest chain	11	27
number of coreference relations	327	1231

*still schemes differences*

# Chain properties in German and Czech

	GERMAN	CZECH
number of coreference chains	225	550
length of coreference chains (mean value per text)	2.45	3.2
longest chain	11	27
number of coreference relations	327	1231
	<b>996</b>	<b>1027</b>

+ repetitions



- zeros



# Chain properties in German and Czech

	GERMAN	CZECH
number of coreference chains	225	550
length of coreference chains (mean value per text)	2.45	3.2
longest chain	11	27
number of coreference relations	327	1231

# Longest chains in German and Czech

## GERMAN

mostly contain

- personal and demonstrative pronouns
- definite nominal phrases

## CZECH

mostly contain

- hyperthematic named entities

*lexical cohesion  
in GECCO*





(1) Úvahy padesát let od SNP

(3) Povstání a vůbec celou antifašistickou rezistenci považují za významnou událost, která podala nefalšovaný obraz Slovenska, řekl v srpnu 1994 v rozhovoru pro slovenský deník Pravda předseda Slovenského svazu protifašistických bojovníků generál Ján Husák.

(4) Také nejvýznamnější slovenští politici od prezidenta přes premiéra až po předsedu parlamentu považují Slovenské národní povstání za pozitivní jev ve slovenských dějinách.

(5) Dnes je už známé, že historie SNP byla překroucena ve prospěch komunistické ideologie.

(6) Po listopadu 1989 sice tento černobílý pohled zmizel, ale vakuum využili převážně exiloví historici a spojili obhajobu Slovenského státu s odsouzením SNP.

(7) Oprostit SNP od nánosu lži a mýtů si kladla za cíl první vědecká konference o SNP, která se uskutečnila letos v dubnu v Donovalech.

(8) S blížícím se 50. výročím SNP dochází na politické scéně i ve slovenském tisku k souboji dvou diametrálně odlišných koncepcí výkladu SNP: na straně jedné se povstání označuje za puč proti vlastní státnosti a na straně druhé je SNP chápáno jako vědomé odmítnutí totality a fašismu.

(9) Podle názoru místopředsedkyně Slovenské národní strany (SNS) Evy Slavkovské přineslo SNP slovenskému národu jen to, že povstal se zbraní v ruce proti fašismu a nepřijal ho trpně jako český národ.

(10) Ø Vzalo mu však mnohem víc: vlastní samostatnost.

(11) Podle komentáře slovenského levicového listu Pravda povstání bránilo hodnoty, na kterých Evropa stála a stojí.

(12) SNS a další národně orientované strany (např. Křesťanská sociální unie, Společnost dr. Jozefa Tisa, Slovenská národní jednota) se přiklánějí k názoru, že šlo o puč proti vlastní státnosti.

(13) Podle představitelů Slovenské národní jednoty Češi v SNP obrali Slováky o vlastní stát a na 40 let jim vnutili čechokomunistický režim.

(14) Strana demokratické levice po přehodnocení faktu, že vedoucí silou SNP byla komunistická strana, se jednoznačně staví za antifašistický a demokratický odkaz povstání.

(15) O postoji Hnutí za demokratické Slovensko (HZDS) k SNP se dá říci, že je přinejmenším neujasněný.

(16) Na loňských oslavách 49. výročí SNP předseda HZDS a bývalý premiér Vladimír Mečiar sice zvěřili.

(17) Ani postoj Ø Křesťanskodemokratického hnutí (KDH) nebyl vždy jasný.

(19) Představitelé KDH však neodmítli dialog a dnes považují povstání za významnou kapitolu slovenské historie.

(20) Slovenské národní povstání politicky organizovala Slovenská národní rada, která vznikla již v září 1943 a zastupovala komunistický a občanský odboj.

(22) Podle Vánoční dohody bylo cílem ozbrojeného boje obnovení Československé republiky.

(24) Během SNP působily na území Slovenska vojenské mise Spojenců, a to sovětská, britská a americká.

(27) Povstání zkomplikovalo plány nacistů na obranu Karpat, připoutalo nemalé nacistické jednotky na Slovensko, rozrušilo komunikační systém v týlu fronty a vytvořilo nebezpečné ohnisko nepokojů.

(28) Ø Posilovalo také odpor v sousedních zemích.

(29) Pro Slovensko Ø mělo význam také politický: Ø zlikvidovalo chápání Slovenska na mezinárodní scéně jako německého vazala.

(1) Úvahy padesát let od **SNP**

(3) **Povstání** a vůbec celou antifašistickou rezistenci považují za významnou událost, **která** podala nefalšovaný obraz Slovenska, řekl v srpnu 1994 v rozhovoru pro slovenský deník Pravda předseda Slovenského svazu protifašistických bojovníků generál Ján Husák.

(4) Také nejvýznamnější slovenští politici od prezidenta přes premiéra až po předsedu parlamentu považují **Slovenské národní povstání** za pozitivní jev ve slovenských dějinách.

(5) Dnes je už známé, že historie **SNP** byla překroucena ve prospěch komunistické ideologie.

(6) Po listopadu 1989 sice tento černobílý pohled zmizel, ale vakuum využili převážně exiloví historici a spojili obhajobu Slovenského státu s odsouzením **SNP**.

(7) Oprostit **SNP** od nánosu lži a mýtů si kladla za cíl první vědecká konference o **SNP, která** se uskutečnila letos v dubnu v Donovalech.

(8) S blížícím se 50. výročím **SNP** dochází na politické scéně i ve slovenském tisku k souboji dvou diametrálně odlišných koncepcí výkladu **SNP**: na straně jedné se **povstání** označuje za **puč** proti vlastní státnosti a na straně druhé je **SNP** chápáno jako vědomé odmítnutí totality a fašismu.

(9) Podle názoru místopředsedkyně Slovenské národní strany (SNS) Evy Slavkovské přineslo **SNP** slovenskému národu jen to, že povstal se zbraní v ruce proti fašismu a nepřijal ho trpně jako český národ.

(10) **Ø** Vzalo mu však mnohem víc: vlastní samostatnost.

(11) Podle komentáře slovenského levicového listu Pravda **povstání** bránilo hodnoty, na kterých Evropa stála a stojí.

(12) SNS a další národně orientované strany (např. Křesťanská sociální unie, Společnost dr. Jozefa Tisa, Slovenská národní jednota) se přiklánějí k názoru, že šlo o **puč** proti vlastní státnosti.

(13) Podle představitelů Slovenské národní jednoty Češi v **SNP** obrali Slováky o vlastní stát a na 40 let jim vnutili čechokomunistický režim.

(14) Strana demokratické levice po přehodnocení faktu, že vedoucí silou **SNP** byla komunistická strana, se jednoznačně staví za antifašistický a demokratický odkaz **povstání**.

(15) O postoji Hnutí za demokratické Slovensko (HZDS) k **SNP** se dá říci, že je přinejmenším neujasněný.

(16) Na loňských oslavách 49. výročí **SNP** předseda HZDS a bývalý premiér Vladimír Mečiar sice zvučeli.

(17) Ani postoj **Ø** Křesťanskodemokratického hnutí (KDH) nebyl vždy jasný.

(19) Představitelé KDH však neodmítli dialog a dnes považují **povstání** za významnou kapitolu slovenské historie.

(20) **Slovenské národní povstání** politicky organizovala Slovenská národní rada, která vznikla již v září 1943 a zastupovala komunistický a občanský odboj.

(22) Podle Vánoční dohody bylo cílem ozbrojeného boje obnovení Československé republiky.

(24) Během **SNP** působily na území Slovenska vojenské mise Spojenců, a to sovětská, britská a americká.

(27) **Povstání** zkomplikovalo plány nacistů na obranu Karpat, připoutalo nemalé nacistické jednotky na Slovensko, rozrušilo komunikační systém v týlu fronty a vytvořilo nebezpečné ohnisko nepokojů.

(28) **Ø** Posilovalo také odpor v sousedních zemích.

(29) Pro Slovensko **Ø** mělo význam také politický: **Ø** zlikvidovalo chápání Slovenska na mezinárodní scéně jako německého vazala.

Das deutsche Modell **<der sozialen Marktwirtschaft>** ist eine ökonomische Erfolgsgeschichte : **<Sie>** verbindet wirtschaftlichen Wohlstand und soziale Gerechtigkeit auf eine Weise, die ihresgleichen sucht . Auch wenn einige **<ihrer Errungenschaften>** der Reform bedürfen - **<sie>** hat schon mehrfach bewiesen, dass **<sie>** die Kraft hat , Umbrüche zu bewältigen. Arbeiten wie die Polen , aber leben wie die Japaner - die Ökonomen der westlichen Welt lächelten milde über die Ideen der östlichen Nachbarn. Nur der französische Wirtschaftsphilosoph Michel Albert verteidigte sie 1991 gänzlich unironisch: "Ist eigentlich bekannt , dass Deutschland von dieser Vorstellung gar nicht so weit entfernt ist ? " Ob eine Wirtschaftsordnung nach deutschem Muster den Umbruch in Polen, Tschechien oder Russland besser gestaltet hätte , lässt sich nun nicht mehr herausfinden. In Deutschland allerdings hat **<die soziale Marktwirtschaft>** schon mehr als einen Umbruch bewältigt und dem Land die langjährige "Exportweltmeisterschaft" eingetragen . **<Sie>** hat wirtschaftlichen Wohlstand und soziale Gerechtigkeit miteinander verbunden, in einer Intensität , die weltweit ihresgleichen sucht. Auch wenn der deutsche Konjunkturmotor zur Zeit eher stottert als brummt: Die ökonomische Erfolgsgeschichte Deutschlands rechtfertigt die Suche nach ihren Erfolgsgeheimnissen . Dass es sich um reichlich deutsche Geheimnisse handelt , zeigt sich schon daran , dass die dazugehörigen Begriffe nicht nur allesamt der deutschen Sprache entstammen , sondern ebenso allesamt praktisch unübersetzbar sind . Als Superstar **<der sozialen Marktwirtschaft>** gilt aus gutem Grund Ludwig Erhard.

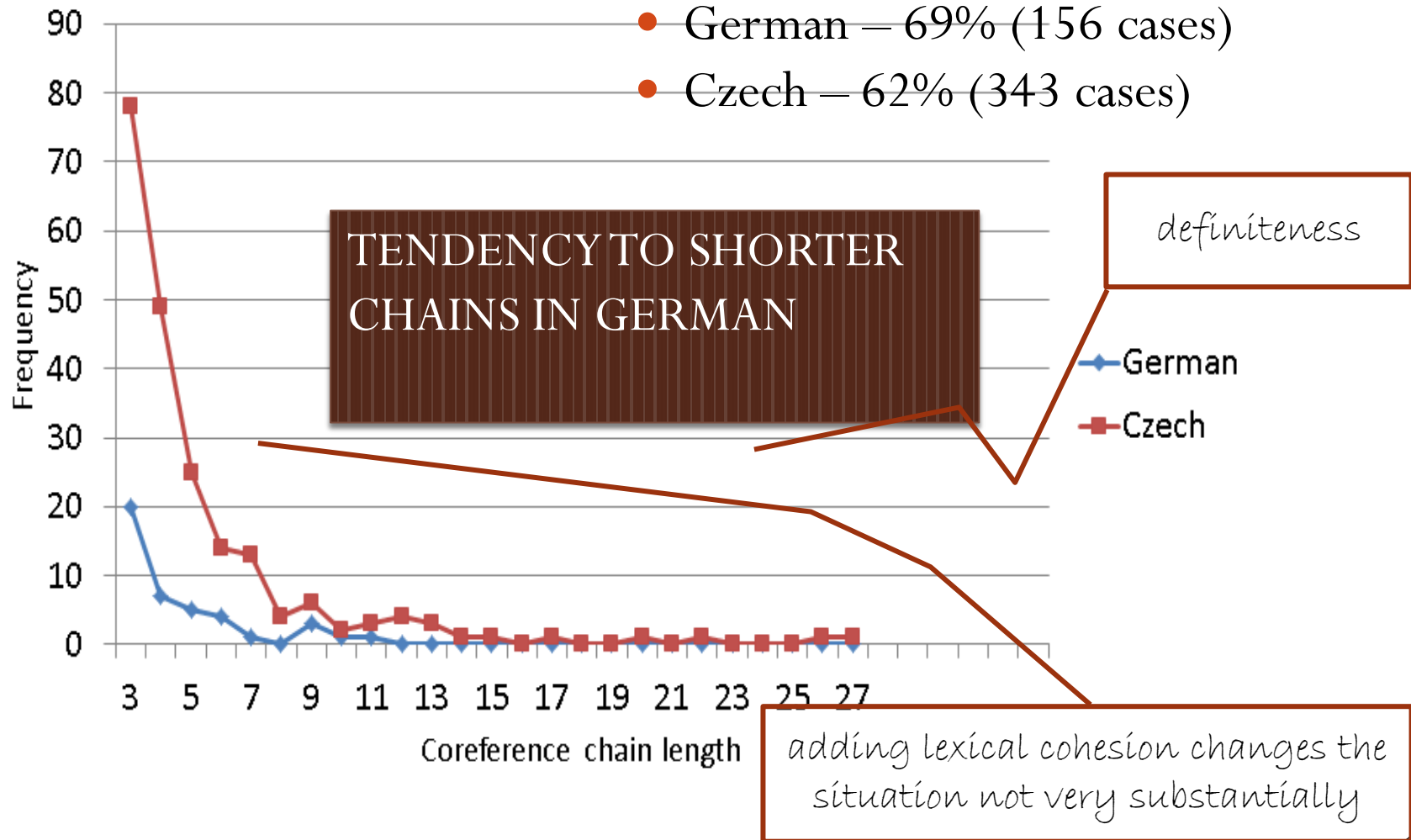
# Frequency of Coreference Chains of Various Length

more than 2-element chains:

Two-element coreference chains:

• German – 69% (156 cases)

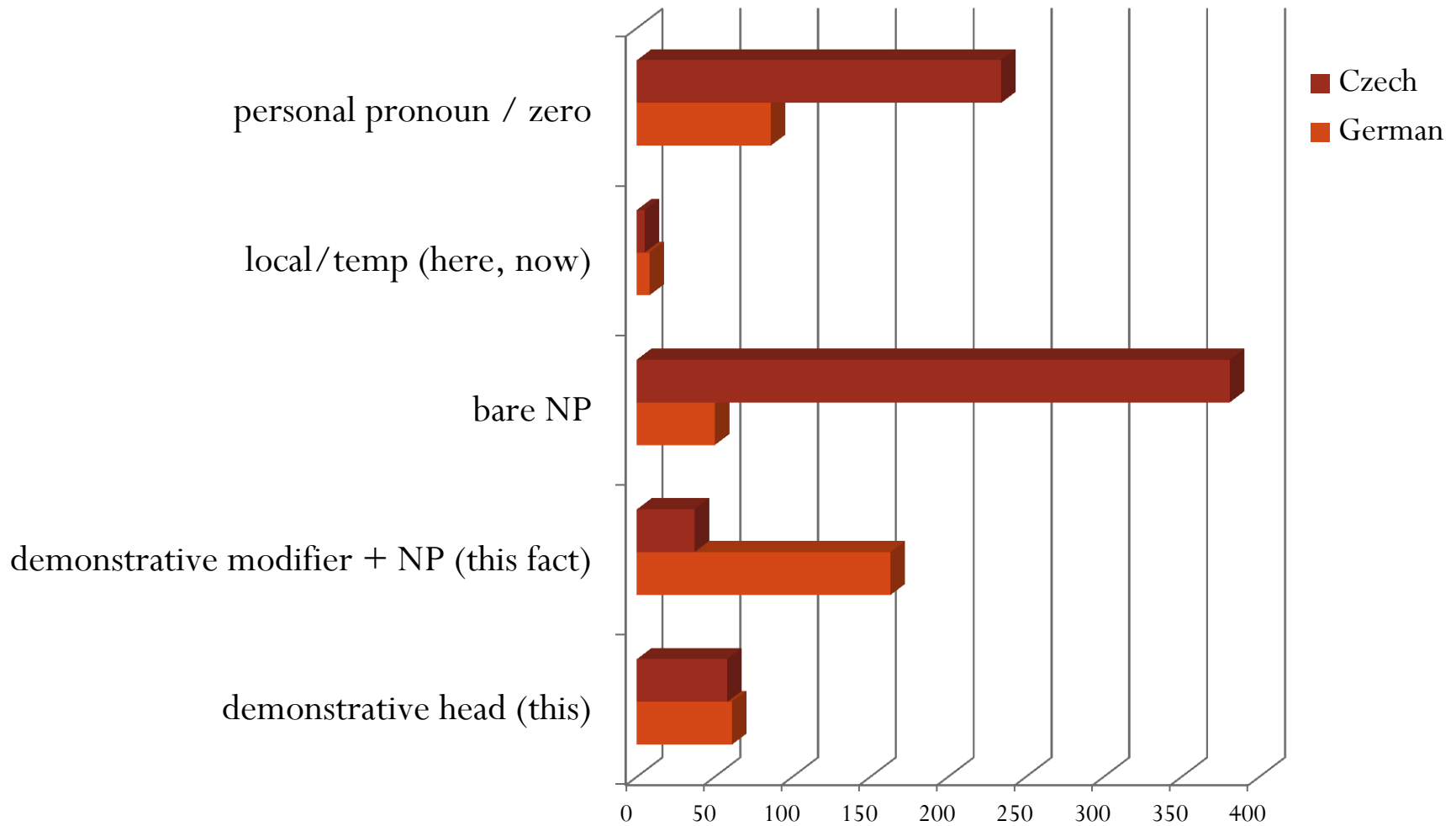
• Czech – 62% (343 cases)



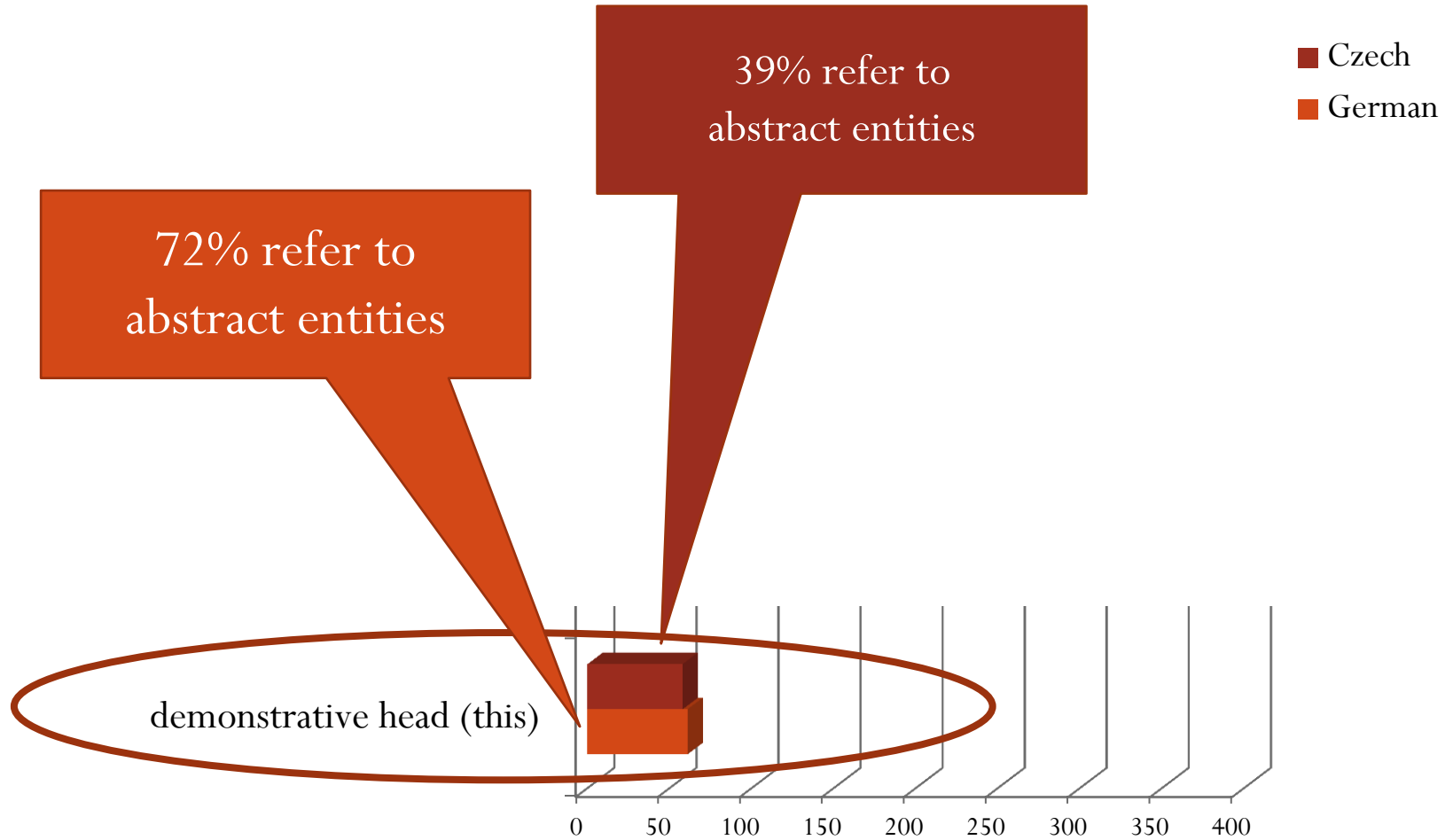
## Results so far...

- comparable results as for the properties of coreference chains
- properties of chains - genre- or register-dependent???
- thematic similarities or differences of the selected texts

# Expressive means of coreference



# Antecedents of Demonstrative Heads



# Abstract Anaphors referring to...

	GERMAN		CZECH	
	absolute	%	absolute	%
to clauses	5	7.35	13	20.63
to sentences	45	66.18	20	31.75
to bigger segments	18	26.47	30	47.62
TOTAL	68	100	63	100



# Expressing means of Abstract Anaphors...

demonstrative head ( <i>dies, dazu/ ten</i> [this])	44	65%	28	44%
demonstrative modifier + NP ( <i>diese Frage/ tato otázka</i> [this question])	16	24%	17	27%
bare NP	0	0%	10	16%
temporal/local ( <i>hier, da, nun/ tam, tady</i> [here, there, now])	3	4%	4	6%
personal pronoun ( <i>es</i> [ $\approx$ <i>it</i> ])/ zero anaphora	3	4%	2	3%
comparative	2	3%	2	3%
total	68	100	63	100

# Logically...

- demonstrative heads (this) – often to clauses

*Cizinci podstatně přispěli k německému hospodářskému a kulturnímu vývoji, proč jejich počet naopak ve statistikách nezdůrazňovat a tím veřejně uznat jejich zásluhy o německou hospodářskou a politickou demokracii?*  
[Foreigners have contributed significantly to the German economic and cultural development, so why not to emphasize their number in statistics, and to acknowledge their merit of the German economic and political democracy by this?]

- (dem.mod.+ )NP (this/ the decline) – often to sentences

*Dnes se tento počet snížil na asi půl milionu, jenže důvodem <poklesu> je především skutečnost, že ten, kdo není zaměstnán déle než rok, již podporu nedostane.* [Today, that number dropped to about half a million, but the reason for the decline is the fact that anyone who is not employed for more than a year, gets no support anymore.]

- dem.mod.+ general noun – often to larger segments

# German: demonstrative head → sentence

Gleichzeitig brauchen wir mindestens eine Verdoppelung des Wohlstands.

Wenn wir die Armutsgegenden der Erde anschauen, weiß jeder sofort, dass dies das Mindeste an moralischer Herausforderung ist.

[At the same time, we need to double the current level of prosperity. One look at the poor regions throughout the world is enough to make anyone realize that this is the most urgent moral challenge we face.]

German: demonstrative modifier + NP  
→ verbal phrase

*Es ist eine der wichtigsten Aufgaben des Staates, die Erhaltung des freien Wettbewerbs sicherzustellen.  
Versagt der Staat auf diesem Felde, dann ist es bald um die soziale Marktwirtschaft geschehen. [Protecting free competition is one of the state's most important tasks. If the state fails in this area, the social market economy will soon be lost].*

German: local → sentence

*Nicht multinationale Unternehmen und Großbanken schaffen die meisten Arbeitsplätze, sondern der Mittelstand. Hier zeigt sich auch die Bedeutung der Berufsausbildung. [The majority of jobs are not created by multinational conglomerates, but by the "Mittelstand", the small and medium-sized enterprises. The importance of vocational training also becomes evident here.]*

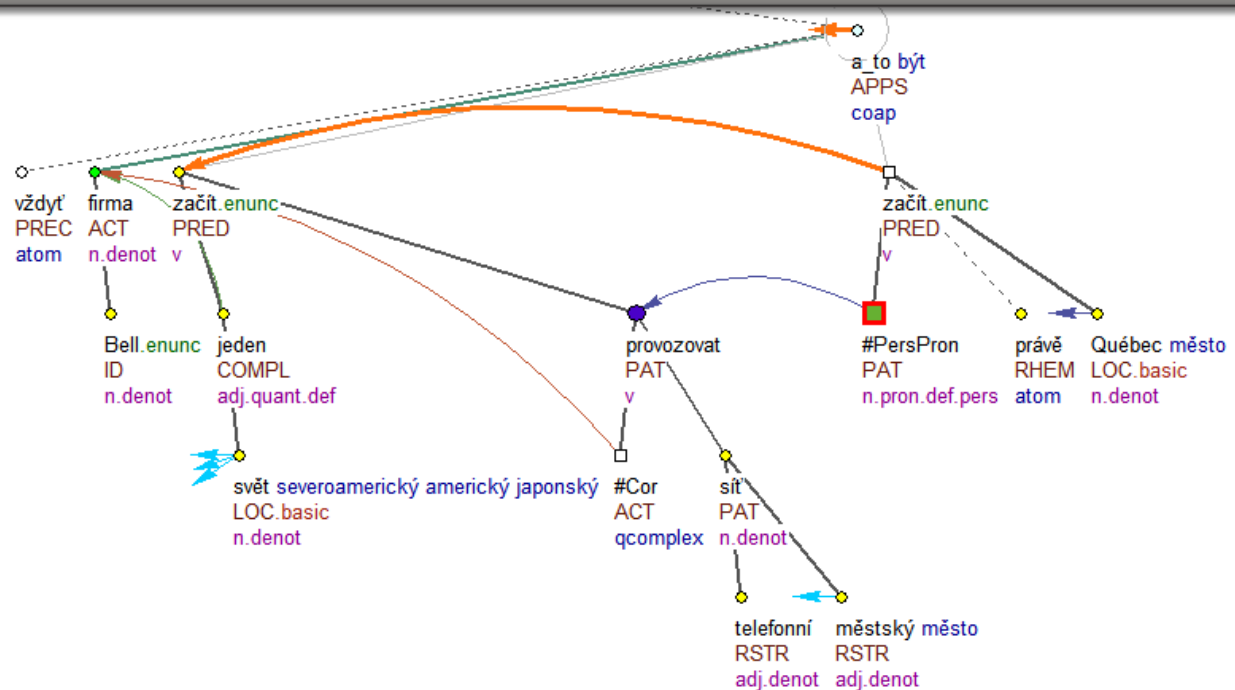
# German: personal pronoun (es) → bigger textual segment

*Jeder hatte sich darin auf irgendeine Weise definiert, nicht zuletzt auch die Vertreter des Kapitals in den marktwirtschaftlich organisierten Demokratien. Für sie war es wichtig, den Beweis zu erbringen, dass die demokratisch organisierte Marktwirtschaft auch sozial dem ineffizienten autoritären Kommunismus weit überlegen war. Es war ein voller Erfolg bis zu dem Punkt, als der OstWest-Konflikt zugunsten des Westens entschieden war.*

*[Everyone defined themselves in some way or another in accordance with this; not least those who represented investments in the market economies underlying the democratic countries . For such people, it was important to produce proof that democratically organized market economies were far superior to inefficient authoritarian communism in social as well as economic terms. It was a "roaring success" until the time came when the East-West conflict was resolved in favour of the West, at which point it was no longer necessary to furnish proof.]*

# Czech: zero anaphora → verbal phrase

*Vždyť firma Bell začala jako první na světě provozovat městskou telefonní síť - a to právě v Québecu. – [Indeed, the Bell company started to operate the urban telephone network first in the world, and it was in Quebec.]*



## Czech: temporal → clause

*Poskytla mi smích už když jsem četl rukopis.*

*Zdeňkovi však tehdy do smíchu nebylo. [It made me laugh when I was reading the manuscript. However, Zdenek, had no intention to laugh then.]*



## Czech: demonstrative head → clause

*Cizinci podstatně přispěli k německému hospodářskému a kulturnímu vývoji, proč jejich počet naopak ve statistikách nezdůrazňovat a tím veřejně uznat jejich zásluhy o německou hospodářskou a politickou demokracii? [Foreigners have contributed significantly to the German economic and cultural development, so why not to emphasize their number in statistics, and to acknowledge their merit of the German economic and political democracy by this?]*

## Czech: bare NP → sentence

*Dnes se tento počet snížil na asi půl milionu.*  
*Důvodem poklesu je především skutečnost, že ten, kdo není zaměstnán déle než rok, již podporu nedostane. [Today, that number dropped to about half a million. However, the reason for the decline is the fact that anyone who is not employed for more than a year, gets no support anymore.]*

## Czech: demonstrative head → bigger segment

V současnosti pracuje ve věznicích České republiky asi padesát procent z 9671 odsouzených. Poslední údaje uvádějí, že v půli srpna bylo ve vězení také 8467 obviněných, celkově tedy 18138 osob. Informoval o tom LN Eduard Vacek. [At present, about fifty percent of the 9,671 inmates work in the prisons of Czech Republic. The latest data show that in mid-August there were charged in jail 8467, a total of 18,138 persons. Eduard Vacek informed LN about this.]

# Conclusion



- technical issues
- differences of typological character (absence of definiteness or pro-drops) *also* have influence on the properties of coreference chains
- the main differences in the length and number of chains (reflecting the topic structuring) are rather genre- or even domain-dependent

# References

- BBN Technologies. Coreference Guidelines for English OntoNotes – Version 6.0. Linguistic Data Consortium. BBN Pronoun Coreference and Entity Type Corpus. 2006.
- Baker, M. (1995), Corpora in translation studies: An overview and some suggestions for future research. *Target* 7(2). 223–243.
- Botley, S. (2006), Indirect anaphora: Testing the limits of corpus-based linguistics, *International Journal of Corpus Linguistics*, 11(1), pp. 73–112.
- Caselli, T., Prodanof I. (2010), Annotating Event Anaphora: A Case Study, *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, Valletta, Malta.
- Dipper S., Zinsmeister H. (2009), Annotating discourse anaphora. *Proceedings of LAW III*, pp. 166–169.
- Francis G. (1994), Labelling discourse: an aspect of nominal group lexical cohesion. Coulthard M. (ed.), *Advances in Written Text Analysis*, London: Routledge, pp. 83–101.
- Fraurud K. (1992), Situation reference: What does 'it' refer to? GAP Working Paper No 24, Fachbereich Informatik, Universität Hamburg.
- Halliday M., Hasan R. (1976), *Cohesion in English*. Longman, London, New York.
- Hansen-Schirra S., Neumann, S., and Steiner, E. (2012), *Cross-linguistic Corpora for the Study of Translations. Insights from the Language Pair English-German*. de Gruyter, Berlin, New York.
- Hedberg N., Jeanette K. G., and Zacharski, R. (2007), Directly and indirectly anaphoric demonstrative and personal pronouns in newspaper articles, *Proceedings of DAARC-2007*, pp. 31–36.
- Kučová L., Hajičová, E. (2004), Coreferential Relations in the Prague Dependency Treebank. *Proceedings of DAARC-2004*, pp. 97–102.
- Lapshinova-Koltunski E., Nedoluzhko A., Kunz K. A. (2015), Across Languages and Genres: Creating a Universal Annotation Scheme for Textual Relations. *Proceedings of the The 9th Linguistic Annotation Workshop (LAW IX 2015) Association for Computational Linguistics, Stroudsburg, PA, USA*, pp. 168-177.
- Lapshinova-Koltunski, E., Kunz, K. (2014), Annotating cohesion for multilingual analysis. *Proceedings of the 10th Joint ACL - ISO Workshop on Interoperable Semantic Annotation*, Reykjavik, Iceland, May. LREC.
- Lapshinova-Koltunski, E., Kunz, K., and Amoia, M. (2012). Compiling a multilingual spoken corpus. In Mello H., Pettorino M., T. R. (ed.), *Proceedings of the VIIth GSCP International Conference: Speech and corpora*, Firenze, Firenze University Press, pp. 79–84.
- Müller Ch. (2008), Fully Automatic Resolution of 'it', 'this', and 'that' in Unrestricted Multi-Party Dialog. Ph.D. thesis, University of Tübingen.
- Navarretta C., Olsen S. (2009), The annotation of pronominal abstract anaphora in Danish texts and dialogues. DAD report 1. Centre for Language Technology, University of Copenhagen.
- Nedoluzhko A. (2011), Extended textual coreference and bridging anaphora [Rozšířená textová koreference a asociální anafora], Prague, Czech Republic, 268 pp., Dec 2011
- Nedoluzhko A., Toldova S., Novák M. (2015), Coreference chains in Czech, English and Russian: Preliminary findings. *Computational Linguistics and Intellectual Technologies*, Vol. 14, No. 21, RGGU, Moscow, Russia, ISSN 2221-7932, pp. 474-486.
- Nedoluzhko, A., Lapshinova-Koltunski, E. (2016), Abstract Coreference in a Multilingual Perspective: a View on Czech and German. In *Proceedings of the CORBON Workshop at NAACL2016*, San Diego, California.
- Pradhan S., Ramshaw L., Weischedel R., MacBride J., Micciulla L. (2007), Unrestricted coreference: Identifying entities and events in OntoNotes. In *Proceedings of the IEEE-ICSC*.
- Recasens M., Martí, A. (2010), AnCoraCO: Coreferentially annotated corpora for Spanish and Catalan. *Language Resources & Evaluation*.
- Recasens, M., Martí, M.-A., Taulé, A.M. (2007), Text as Scene: Discourse Deixis and Bridging Relations. *Procesamiento del Lenguaje Natural*, 39.
- Zinsmeister, H., Dipper, S., Seiss, M. (2012). Abstract pronominal anaphors and label nouns in german and english: selected case studies and quantitative investigations. *Translation: Computation, Corpora, Cognition* 2(1).
- Zikánová Š., Hajičová E., Hladká B., Jínová P., Mírovský J., Nedoluzhko A., Poláková L., Rysová K., Rysová M., Václ J. (2015), Discourse and Coherence. From the Sentence Structure to Relations in Text. *ÚFAL*, Prague, Czech Republic, 274 pp.