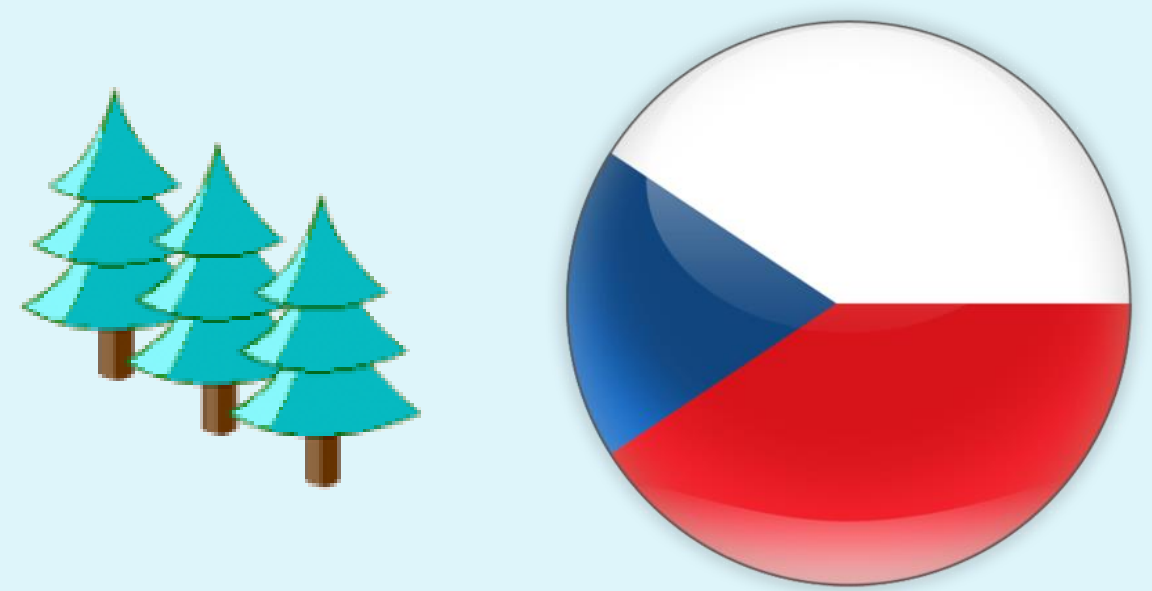




ABSTRACT COREFERENCE IN A MULTILINGUAL PERSPECTIVE: A VIEW ON CZECH AND GERMAN



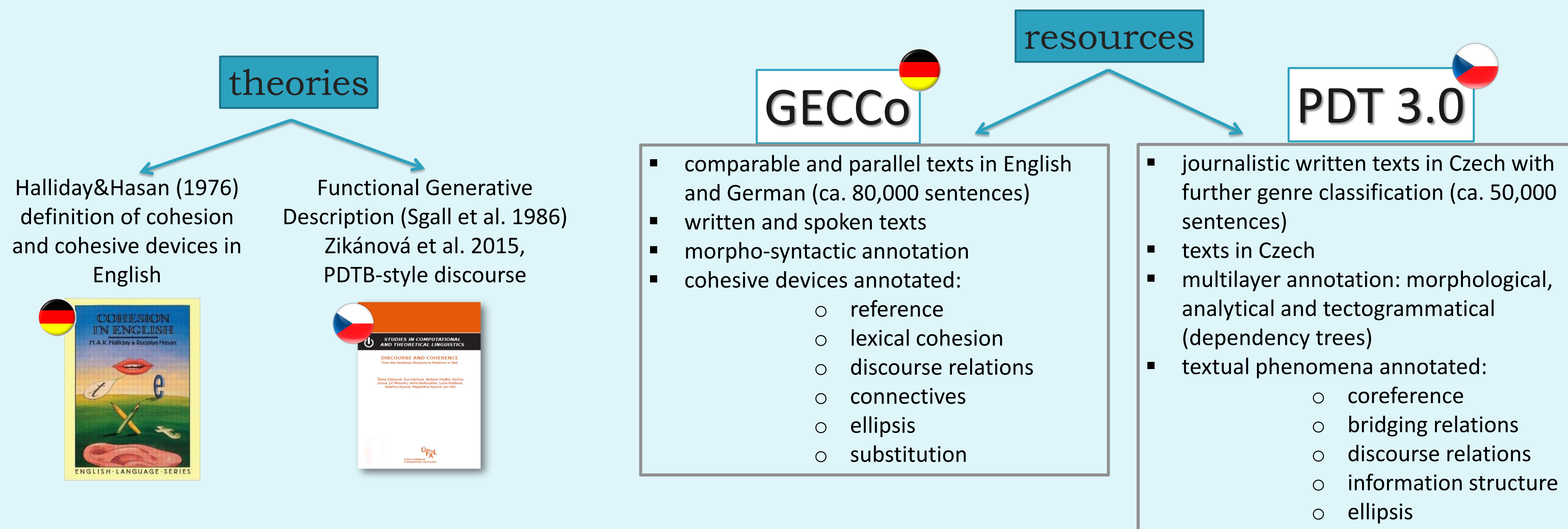
Ekaterina Lapshinova-Koltunski (e.lapshinova@mx.uni-saarland.de, Saarland University), Anna Nedoluzhko (nedoluzko@ufal.mff.cuni.cz, Charles University in Prague)

AIMS AND MOTIVATION

Abstract coreference (= abstract anaphora, discourse anaphora, event anaphora, situational deixis, discourse deixis, reference to events, etc.) in our data (GECCo + PDT):

- ❖ verbal phrases
- ❖ finite clauses and sentences
- ❖ larger text passages and discontinuous strings

- ❖ compare languages within the area of abstract anaphora: Are there any differences between German and Czech with respect to how abstract coreference is expressed?
- ❖ 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



GERMANIC AND SLAVIC LANGUAGES, METHODS AND DATA

	Czech	German
	☐ morphologically rich	☐ reduced morphology
	☐ less isolating	☐ more isolating
	☐ flexible word order	☐ less flexible word order
	☐ no definite article	☐ definite article
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
Total: ALL GERMAN	645	12243

RESULTING SCHEMA APPLIED			
IDENTITY	featID	Czech	German
	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
NON-IDENTITY	id5	coreference with local and temporal adverbs	coreference with local and temporal adverbs
	nonid1	relations of MERONYMY	relations of MERONYMY
	nonid2	bridging CONTRAST	particular comparative reference and antonyms
DISCOURSE RELATIONS	temp	temporal	temporal
	cont	contingency	causal
	comp	comparison (contrast)	adversative
	expan	expansion	additive
ELLIPSIS	ellipsis	textual ellipsis	cohesive ellipsis



id	topics	sent	tokens
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
Total: ALL CZECH		628	11399

ANALYSIS

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
Modified number of coreference relations	996	1027

- zeros

+ repetitions

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].

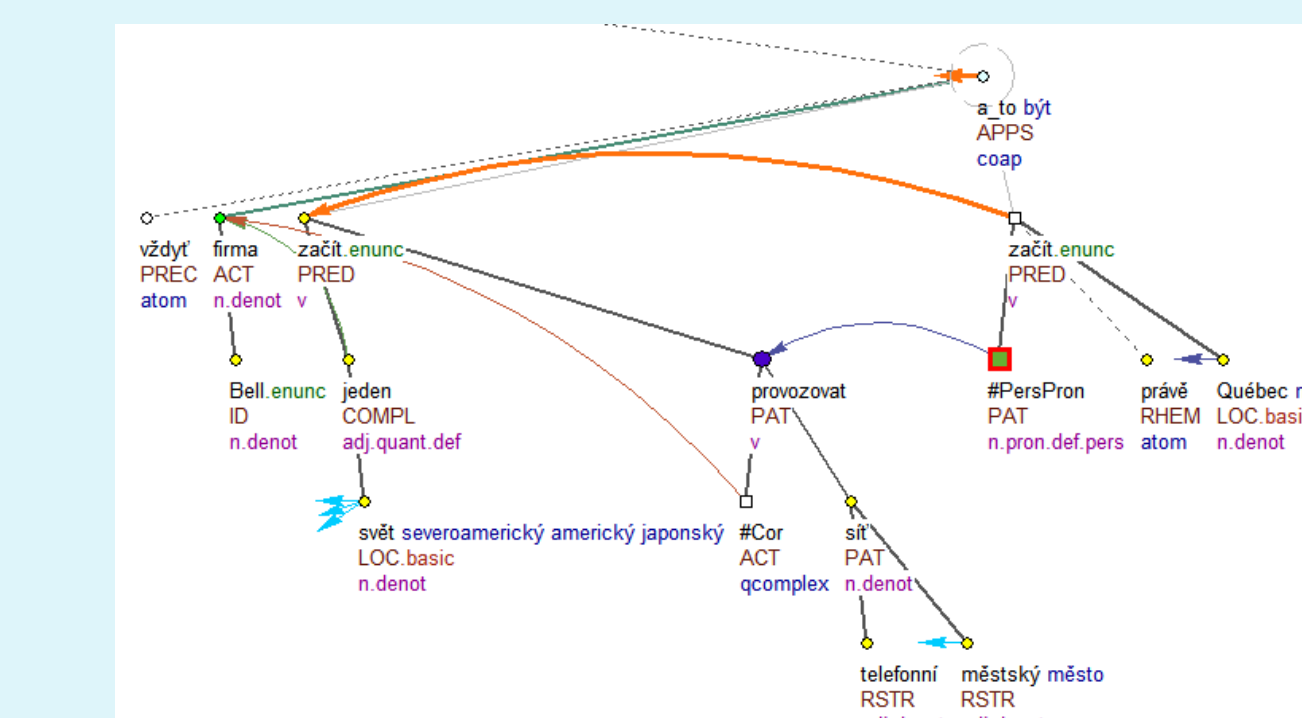
German: local → sentence

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 Ost-West-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.]

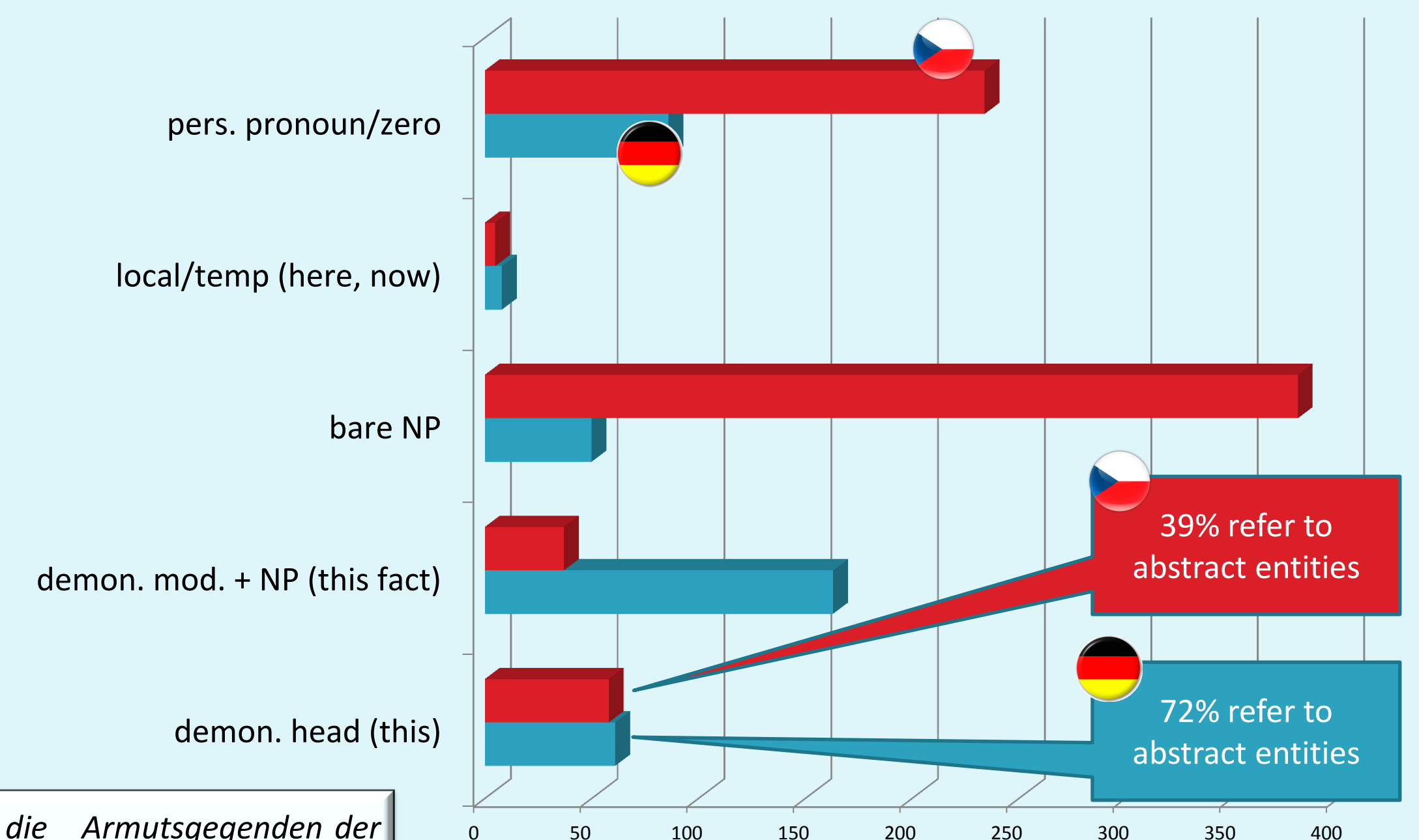
	GERMAN	CZECH
demonstrative head (<i>dies, dazu/ ten</i> [this])	44 65%	28 44%
demonstrative modifier + NP (<i>diese Frage/ tato otázka</i> [this question])	16 24%	17 27%
bare NP	0 0%	10 16%
temporal/local (<i>hier, da, nun/ tam, tady</i> [here, there, now])	3 4%	4 6%
personal pronoun (<i>es</i> [≈ <i>it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



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 was the first in the world to start an urban telephone network, which was in Québec.]

Distribution of anaphors in German and Czech



	abs.	%	abs.	%
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

Czech: temporal → clause

Poskytl 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 more support.]

Czech: demonstrative head → bigger textual 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 8467 such prisoners in jail, a total of 18,138 persons. Eduard Vacek provided LN with this [lit. this information].