

FROM MONOLINGUAL ANNOTATIONS TOWARDS CROSS-LINGUAL RESOURCES: AN INTEROPERABLE APPROACH TO THE ANALYSIS OF DISCOURSE

PDiT

Ekaterina Lapshinova-Koltunski*, Anna Nedoluzhko**, Kerstin Kunz***

Saarland University*, Charles University in Prague**, University of Heidelberg***

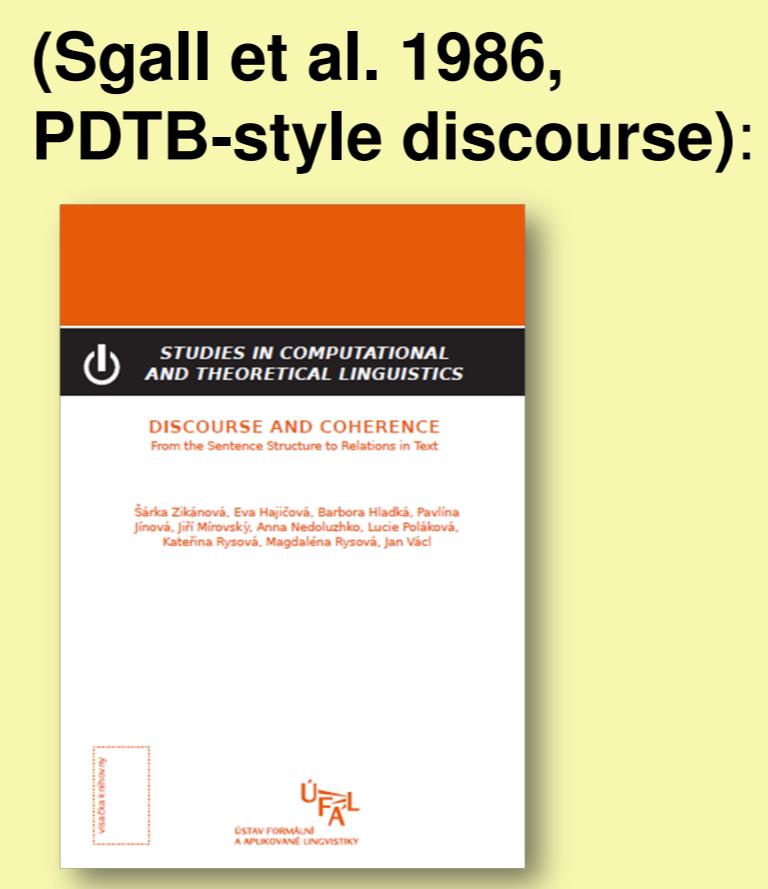
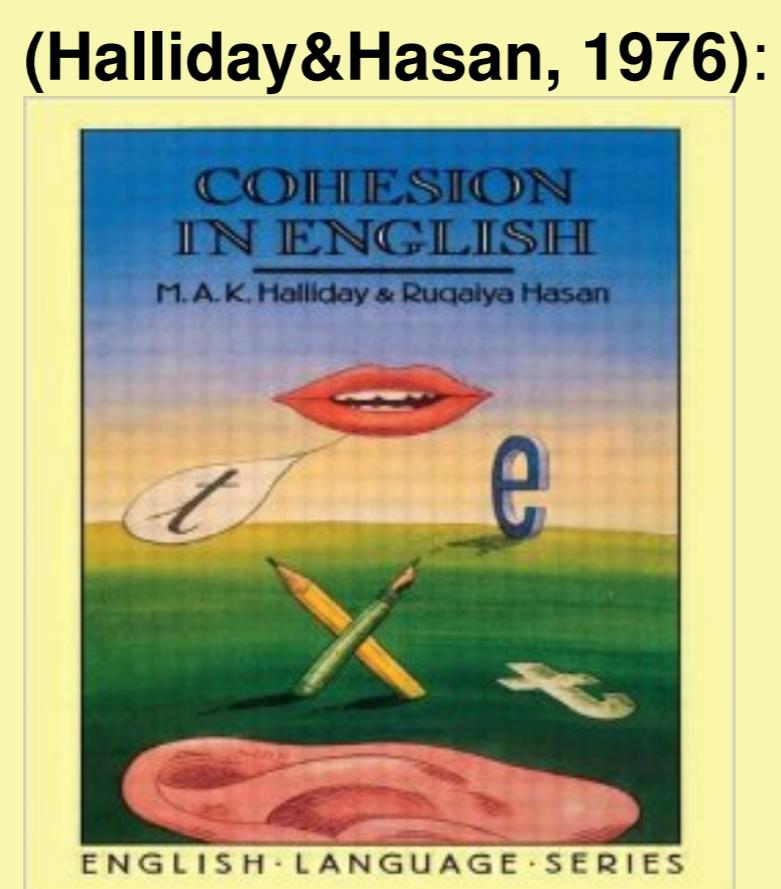
e.lapshinova@mx.uni-saarland.de, nedoluzko@ufal.mff.cuni.cz, kerstin.kunz@iued.uni-heidelberg.de

AIMS AND MOTIVATION

LANGUAGE CONTRASTS



MONOLINGUAL FRAMEWORKS



FURTHER UNIFICATION ANALYSES

presentations

Lapshinova et al. (2015), Nedoluzhko&Lapshinova (2016)
DiSpoL Dialogue
Lapshinova et al. (2015), Nedoluzhko&Lapshinova (2016),
1st Action Conference NAACL6-CORBON
Nedoluzhko et al. (2015), Lapshinova et al (2016),
1st Action Conference LREC
Lapshinova et al. (2015), NAACL-LAW

publications

Nedoluzhko&Lapshinova (2016)
Dialogue
Nedoluzhko&Lapshinova (2016),
NAACL6-CORBON
Lapshinova et al. (2015), NAACL-LAW

GERMANIC AND SLAVIC LANGUAGES, CONTRASTS AND RELATED WORK

CZECH	GERMAN
→ morphologically rich	→ reduced morphology
→ less isolating	→ more isolating
→ flexible word order	→ less flexible word order



creation of discourse



RELATED WORK

Šticha (2003)
Engel (1999)
Komárek (1994)
Novak&Nedoluzhko (2015)
Nedoluzhko et al. (2015)

METHODS AND DATA

SCHEME

	featID	Czech	German
IDENTITY	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
NON-IDENTITY	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

id3 with anaphors realised by nouns

GO: Staatsträger können Gewerkschaften kaum sein. Auch wenn... Ludwig Erhard von den Gewerkschaften nicht viel hielt... [Greater loyalty to the state can hardly be expected of a trade union. Despite the fact that... Ludwig Erhard did not think much of the trade unions...]
CZ: Zaím se posunuje stále více za Prahu... Po dálnici bychom se měli svézt z Prahy až do Českých Budějovic... [So far, people are moving away from Prague... [Highways should take us from Prague all the way to České Budějovice...]

lang	sent	token
GO	736	14930
CZ	763	11769

id1 with pronouns referring to nominal antecedents

GO: Als Superstar der sozialen Marktwirtschaft gilt aus gutem Grund Ludwig Erhard. Er hatte... in den 50er Jahren... die produktiven Kräfte der Unternehmen entfesselt und daraus ein Wirtschaftswunder gezaubert... [Ludwig Erhard is regarded as the superstar of the social market economy, and for good reasons. ...in the nineteen-fifties..., he had unleashed the productive forces of business and in this way conjured up an economic miracle...]
CZ: Ta přijala strategii Bílého domu v domnění, že je to nejjistější cesta k vítězství. [She endorsed the White House strategy, believing it to be the surest way to victory.]

nonid1 with meronymy relations

GO: ...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 danger is much more real, however: that the Germans...]
CZ: 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.]

id2 with abstract anaphors

GO: 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].
CZ: Cizinci podstatně přispěli k německému hospodářskému a kulturnímu vývoji, procíž 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?]

comp with discourse relations of comparison/contrast

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

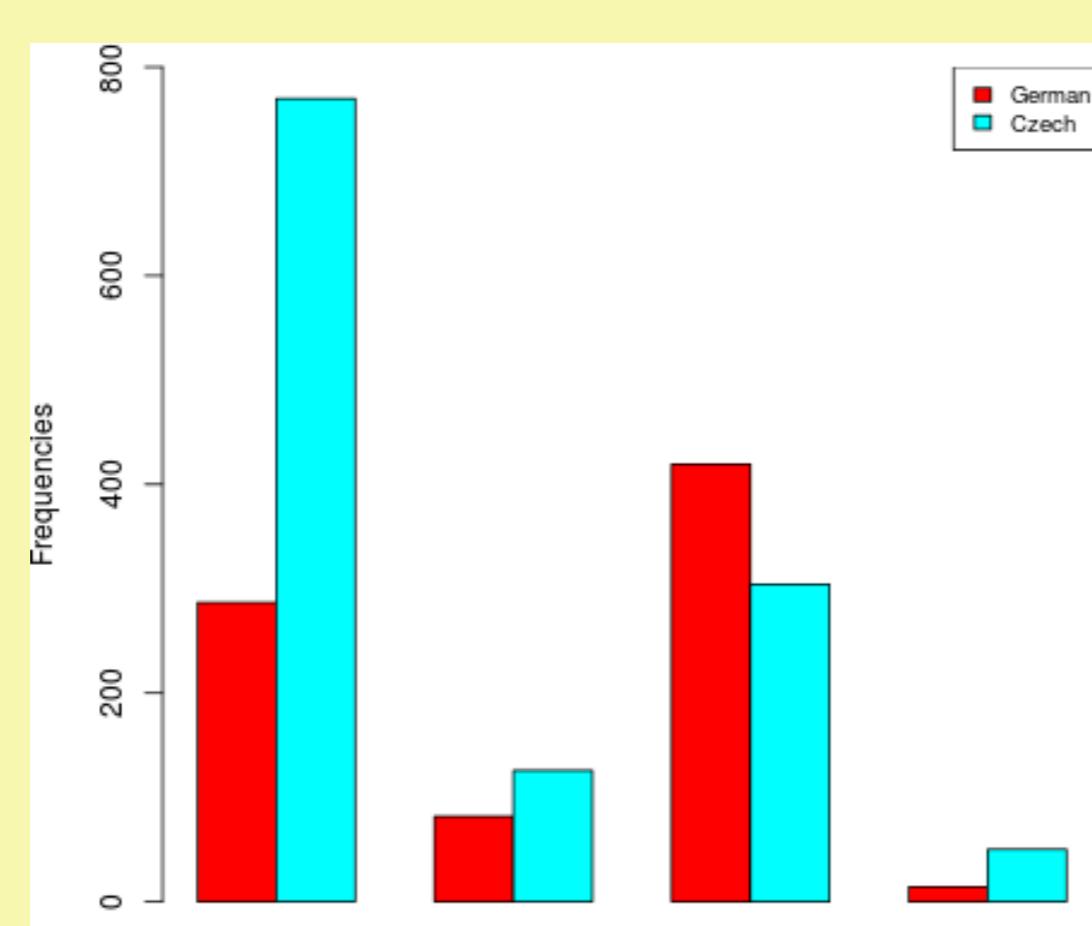
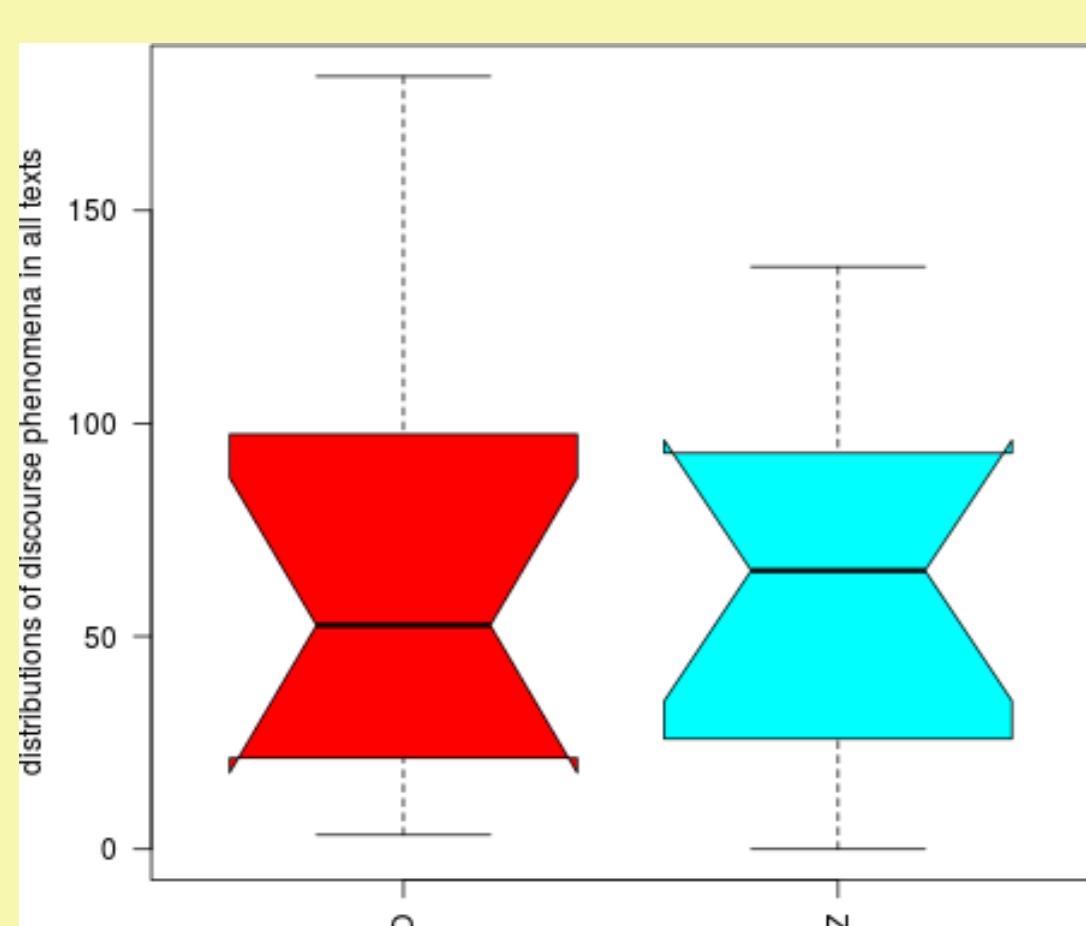
nonid2 with contrast relations

GO: Dazu gehören zum Beispiel die Halbierung der Energie- und Rohstoffintensität bis 2020 gegenüber 1990 (bzw. 1994) 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 (or 1994) levels and doubling the percentage of renewable energy used as part of total energy consumption by 2010.]
CZ: 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 pěsto 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

GO: All das ist eine kleine Revolution. Die grösitere [...] ist diese: [But there is also a bigger [revolution], and it is this:]
CZ: Klienti pojíštoven, které ukončí svou činnost, se automaticky vrátí k Všeobecné [...]. [Clients of insurance companies which shut down will automatically return to the General [one].]

ANALYSES AND RESULTS



SUBCATEGORIES

featID	id1	id 2	id3	id4	id5	nonid1	nonid2	temp	cont	comp	expan	ellipsis
German	88.41	38.18	144.68	3.35	12.06	52.91	28.80	106.50	52.24	79.04	181.51	14.07
Czech	97.71	64.58	597.33	0.00	10.20	88.37	37.39	14.44	66.28	86.67	136.80	50.13

CONCLUSIONS

- cross-lingual analysis of discourse phenomena with resources from different frameworks
- insights into differences between German and Czech
- interoperable schemes permit multilingual analyses of resources from different approaches
- saves time and effort as no compilation of additional resources is required
- limitations the dataset: author- or source dependent variation, size
- differences in the conceptualisations not languages?