



# Multilingual Database – Obstacles and Opportunities

---

Thomas Schmidt, Project Zb



# Multilingual Database – Obstacles and Opportunities

---

- Multilingual database
  - Diversity of multilingual data
  - Points of comparison
- Opportunities and Obstacles
  - Reasons to share data
  - Reasons not to share data

# Data of German/Turkish and German/Portuguese bilinguals in projects E5 and E2

Sel [v]	We nn du	Spanisch willst, musst	Du	Spanisch lernen oder Fransösüs	oder wenn du	((1s)) Dings...				
Sel [tl]	COMP	PRON2SGSpanish	want:PRES:2SGmust:PRES:2SGPron2SGSpanish	learn:INF	CONJ	French	CONJ	COMP	PRON2SG	thing
Sel [ts]	When you want Spanish, you have to learn Spanish or French				or when you ((1s)) how do you call it...					
Yil [v]	Nasil?								Benimle lütfen	
Yil [ts]	How?								Would you mind	
Sel [k]	<i>für: Französisch</i>									

	CHILD	INTERVIEWER	COMMENT
1	<b>hier du mußt hier das hinstell'n</b> U D M U D V		Kind zeigt Interviewerin, wo sie die Spielfiguren hinstellen muss.
2	<b>gib mir meine roten erstmal dann bin ich dein freund</b> V D D A U U I D D N		
3	<b>dann stell' ich die hin aber</b> U V D D U U	da sind deine roten	Interviewer und Y lachen, Interviewer gibt Kind die roten Spielfiguren
4	<b>mußt du so drehn o.k.</b> M D U V U		Kind zeigt Interviewer, wie man das Krokodil drehen kann
7	<b>wenn du hier machst dann mußt du auf den krokodil dreh'n</b> C D U V U M D P D N V		Kind zeigt auf die Würfelseite mit dem Krokodilkopf

# Written (K4, H1, H3, H4, H5) vs. Spoken (K1, K2, K5, E1, E2, E3, E4, E5)

---

## Written

Paragraphs

Sentences

Titles

Authors

## Spoken

Turns

Utterances

Speakers

Overlap

Pronunciation

Hesitation Phenomena

Pauses

[...]

# Different languages / different writing systems

## Latin alphabet(s)

ÄÅÆ Ø å æ ø

Scandinavian (K5)

Â Ç Ğ İ Ì Ö Ş Ş Û Ü â ç ğ î ö ş ş û ü

Turkish (E5)

Ç Ê Ê Í Ì Ò Ó Ô Õ Ú Û à á â ã ç é ê í ò ó ô õ ú ü

Portuguese (E2)

## Non-Latin alphabets

α β γ Ω δ ε ζ η θ ω ψ φ ξ π

Greek (H4)

ɔ ʒ ʧ ʦ ʎ ʌ ʌ ʌ ʒ ɛ

IPA (E3)

## Non-Alphabets

天気, 辞書, 地図, サイト案内,

Japanese (K1)



# Different research goals

---

- Discourse Analysis
  - Turns & Utterances (Overlap)
  - Spoken language phenomena
  - Non-verbal behaviour
- Syntactic Analysis
  - Sentences (IPs, CPs)
  - Phrase structure
  - Inflectional endings
- Phonology / Phonetics
  - Syllables and Phonemes
  - Voice onset time

# Different software

Speaker	Turkish Text	English Text
Ays		
TS-Ays		
Mer	tdi©imden sonra ağıd/bu zaten. • Bulduk...	H? Pizza/
TS-Mer	n bin, wurde es erst eröffnet.	Wir haben es gefunden... Hin? Pizza/ Pizza
Mut	(Pizza) Napoli mi var kar/snda?	Pizza Napoli mi var kar/snda?
TS-Mut		Befindet sich (Pizza) Napoli gegenüber? Befindet sich Pizza Napoli gegenüber?
Sev		
TS-Sev		

syncWriter  
(K1, K2, E5)

HIAT-DOS  
(K5)



# Points of comparison

---

Multilingual data (all projects)

Transcription data (K1, K2, K5, E1, E2, E3, E4 and E5)

Written data (K4, H1, H3, H4, H5)

Historical texts (H-projects)

Acquisition data (E-projects)

German/Romance bilinguals (E1, E2 and E3)

Turkish/German bilinguals (E4 and E5)

Indo-European/Non-Indo-European bilinguals (E2, E4  
and E5)

HIAT data (K1, K2, K5, E5)

English texts (K4 and H5)

"Expert discourse" (K1, K2, K4, K5)



# Reasons to share data

---

- Understand other people's analyses
- Verify or falsify other people's analyses
- Test own hypotheses on other people's data
- Use other people's tools
- Enlarge empirical basis (Representativeness)
- Keep data usable (exchange in time)



# Reasons not to share data

---

Linguistic diversity  
Theoretical diversity  
Technical diversity  
Legal aspects  
Scientific competition  
Priorities: time and effort