

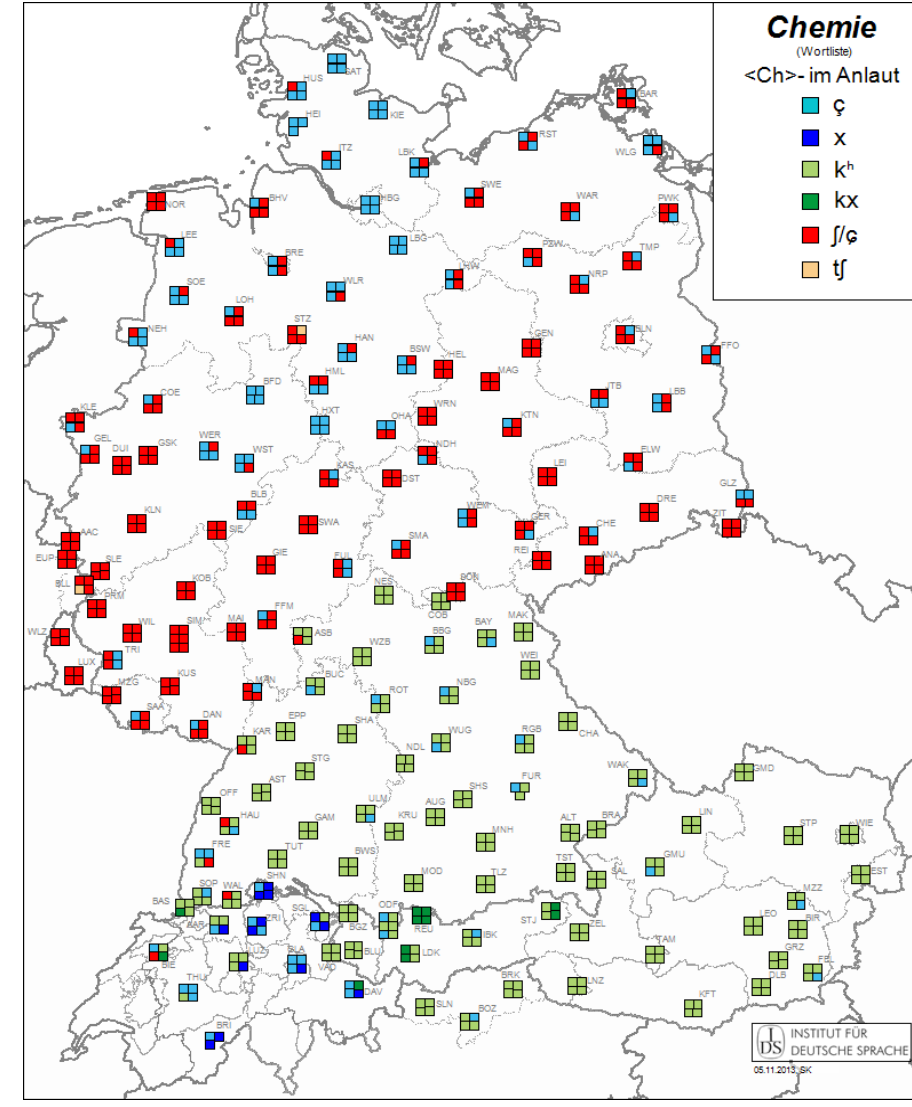
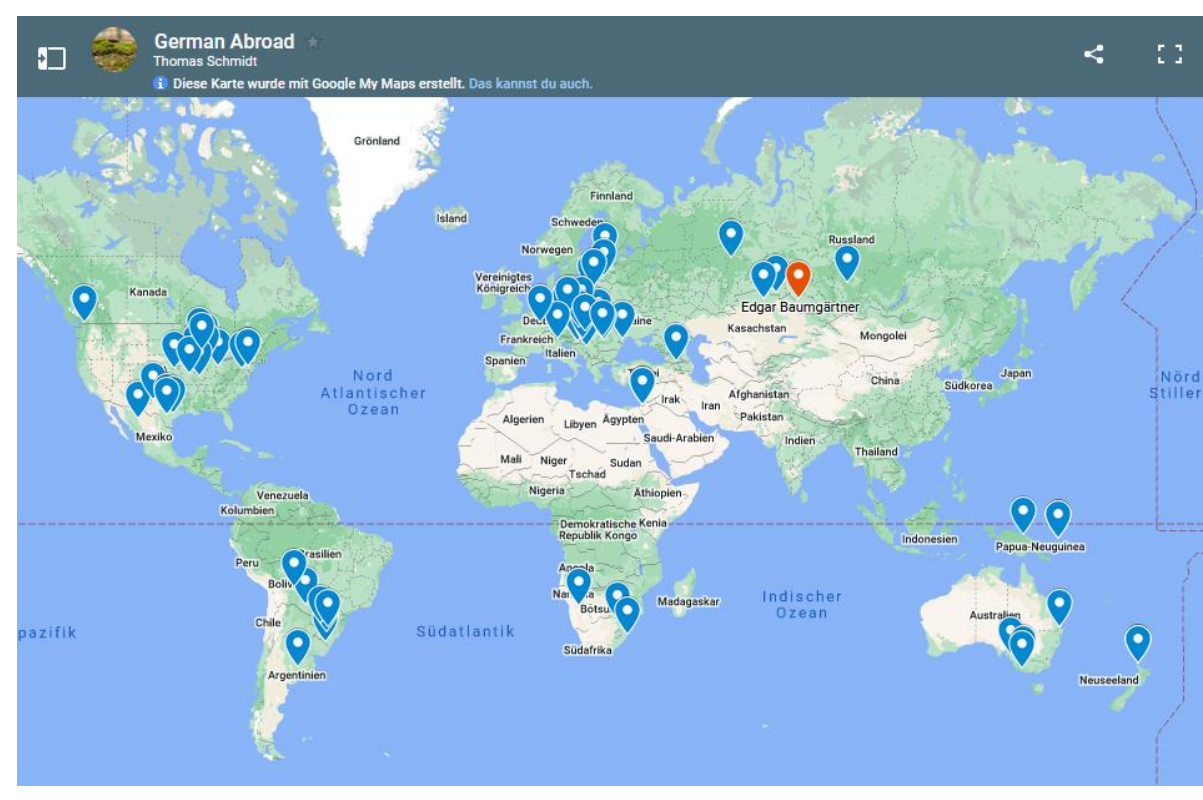
Evaluating Workflows for Creating Orthographic Transcripts for Oral Corpora by Transcribing from Scratch or Correcting ASR-Output

Jan Gorisch¹ & Thomas Schmidt²
¹Leibniz-Institute for the German Language, Mannheim
²Linguisticbits.de, Bingen

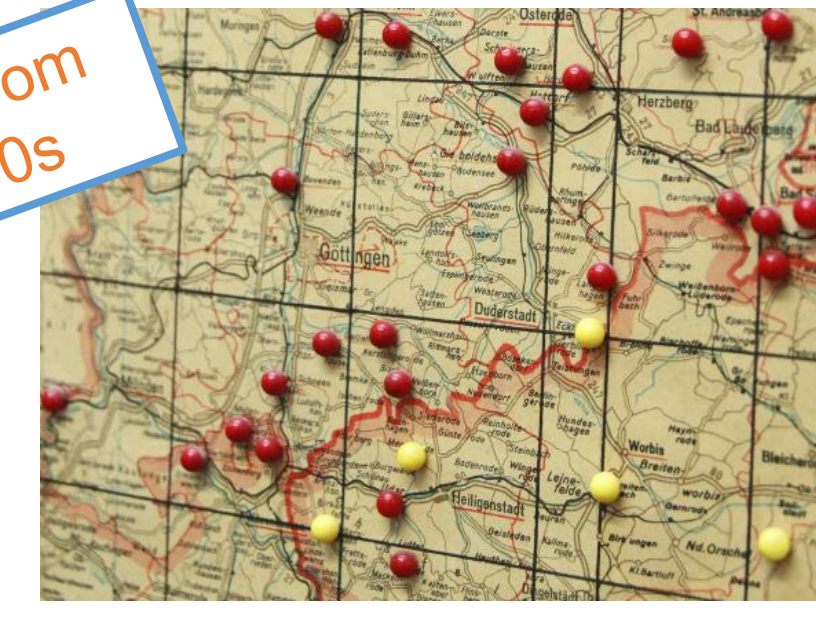
gorisch@ids-mannheim.de, thomas@linguisticbits.de

The AGD* hosts...

- Conversation Corpora
- Variation Corpora
- Extra-territorial Varieties



also historic data from the 1950s and 60s

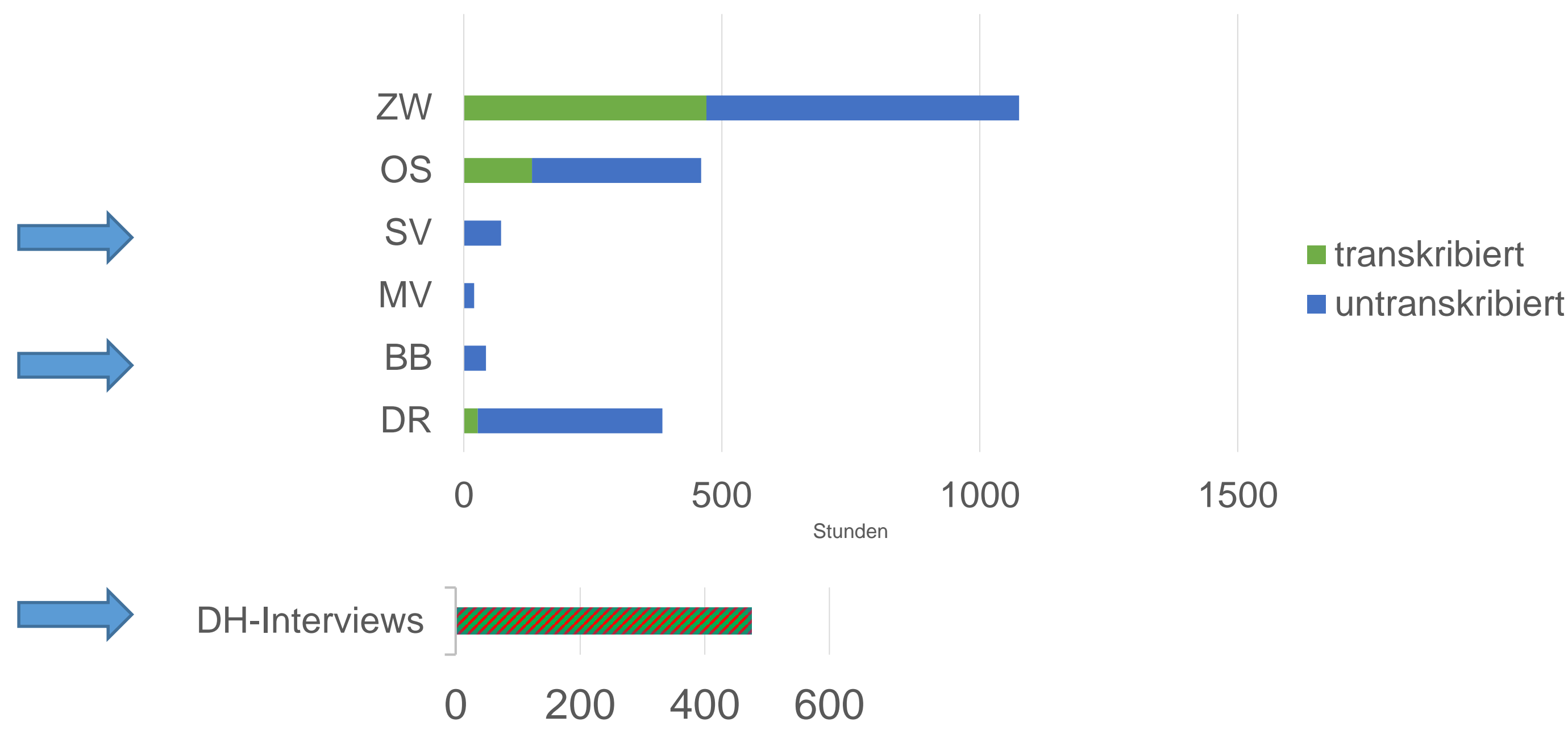


```

Sprecher  Ort  Koordinaten  Geschl.  Geografische_Länge
33  <Sprecher>  <Ort>  <Koordinaten>  <Geschl.>  <Geografische_Länge>
34  <Sprecher>  <Ort>  <Koordinaten>  <Geschl.>  <Geografische_Länge>
35  <Sprecher>  <Ort>  <Koordinaten>  <Geschl.>  <Geografische_Länge>
36  <Sprecher>  <Ort>  <Koordinaten>  <Geschl.>  <Geografische_Länge>
37  <Sprecher>  <Ort>  <Koordinaten>  <Geschl.>  <Geografische_Länge>
38  <Sprecher>  <Ort>  <Koordinaten>  <Geschl.>  <Geografische_Länge>
39  <Sprecher>  <Ort>  <Koordinaten>  <Geschl.>  <Geografische_Länge>
40  <Sprecher>  <Ort>  <Koordinaten>  <Geschl.>  <Geografische_Länge>
41  <Sprecher>  <Ort>  <Koordinaten>  <Geschl.>  <Geografische_Länge>
42  <Sprecher>  <Ort>  <Koordinaten>  <Geschl.>  <Geografische_Länge>
43  <Sprecher>  <Ort>  <Koordinaten>  <Geschl.>  <Geografische_Länge>
44  <Sprecher>  <Ort>  <Koordinaten>  <Geschl.>  <Geografische_Länge>
45  <Sprecher>  <Ort>  <Koordinaten>  <Geschl.>  <Geografische_Länge>
46  <Sprecher>  <Ort>  <Koordinaten>  <Geschl.>  <Geografische_Länge>
47  <Sprecher>  <Ort>  <Koordinaten>  <Geschl.>  <Geografische_Länge>
48  <Sprecher>  <Ort>  <Koordinaten>  <Geschl.>  <Geografische_Länge>
49  <Sprecher>  <Ort>  <Koordinaten>  <Geschl.>  <Geografische_Länge>
50  <Sprecher>  <Ort>  <Koordinaten>  <Geschl.>  <Geografische_Länge>
51  <Sprecher>  <Ort>  <Koordinaten>  <Geschl.>  <Geografische_Länge>
52  <Sprecher>  <Ort>  <Koordinaten>  <Geschl.>  <Geografische_Länge>
53  <Sprecher>  <Ort>  <Koordinaten>  <Geschl.>  <Geografische_Länge>
54  <Sprecher>  <Ort>  <Koordinaten>  <Geschl.>  <Geografische_Länge>
55  <Sprecher>  <Ort>  <Koordinaten>  <Geschl.>  <Geografische_Länge>
56  <Sprecher>  <Ort>  <Koordinaten>  <Geschl.>  <Geografische_Länge>
57  <Sprecher>  <Ort>  <Koordinaten>  <Geschl.>  <Geografische_Länge>
58  <Sprecher>  <Ort>  <Koordinaten>  <Geschl.>  <Geografische_Länge>
59  <Sprecher>  <Ort>  <Koordinaten>  <Geschl.>  <Geografische_Länge>
60  <Sprecher>  <Ort>  <Koordinaten>  <Geschl.>  <Geografische_Länge>
61  <Sprecher>  <Ort>  <Koordinaten>  <Geschl.>  <Geografische_Länge>
62  <Sprecher>  <Ort>  <Koordinaten>  <Geschl.>  <Geografische_Länge>
63  <Sprecher>  <Ort>  <Koordinaten>  <Geschl.>  <Geografische_Länge>
64  <Sprecher>  <Ort>  <Koordinaten>  <Geschl.>  <Geografische_Länge>
65  <Sprecher>  <Ort>  <Koordinaten>  <Geschl.>  <Geografische_Länge>
66  <Sprecher>  <Ort>  <Koordinaten>  <Geschl.>  <Geografische_Länge>
67  <Sprecher>  <Ort>  <Koordinaten>  <Geschl.>  <Geografische_Länge>
68  <Sprecher>  <Ort>  <Koordinaten>  <Geschl.>  <Geografische_Länge>
69  <Sprecher>  <Ort>  <Koordinaten>  <Geschl.>  <Geografische_Länge>
70  <Sprecher>  <Ort>  <Koordinaten>  <Geschl.>  <Geografische_Länge>
71  <Sprecher>  <Ort>  <Koordinaten>  <Geschl.>  <Geografische_Länge>
72  <Sprecher>  <Ort>  <Koordinaten>  <Geschl.>  <Geografische_Länge>
73  <Sprecher>  <Ort>  <Koordinaten>  <Geschl.>  <Geografische_Länge>
74  <Sprecher>  <Ort>  <Koordinaten>  <Geschl.>  <Geografische_Länge>
75  <Sprecher>  <Ort>  <Koordinaten>  <Geschl.>  <Geografische_Länge>
76  <Sprecher>  <Ort>  <Koordinaten>  <Geschl.>  <Geografische_Länge>
77  <Sprecher>  <Ort>  <Koordinaten>  <Geschl.>  <Geografische_Länge>
78  <Sprecher>  <Ort>  <Koordinaten>  <Geschl.>  <Geografische_Länge>
79  <Sprecher>  <Ort>  <Koordinaten>  <Geschl.>  <Geografische_Länge>
80  <Sprecher>  <Ort>  <Koordinaten>  <Geschl.>  <Geografische_Länge>
81  <Sprecher>  <Ort>  <Koordinaten>  <Geschl.>  <Geografische_Länge>
82  <Sprecher>  <Ort>  <Koordinaten>  <Geschl.>  <Geografische_Länge>
83  <Sprecher>  <Ort>  <Koordinaten>  <Geschl.>  <Geografische_Länge>
84  <Sprecher>  <Ort>  <Koordinaten>  <Geschl.>  <Geografische_Länge>
85  <Sprecher>  <Ort>  <Koordinaten>  <Geschl.>  <Geografische_Länge>
86  <Sprecher>  <Ort>  <Koordinaten>  <Geschl.>  <Geografische_Länge>
87  <Sprecher>  <Ort>  <Koordinaten>  <Geschl.>  <Geografische_Länge>
88  <Sprecher>  <Ort>  <Koordinaten>  <Geschl.>  <Geografische_Länge>
89  <Sprecher>  <Ort>  <Koordinaten>  <Geschl.>  <Geografische_Länge>
90  <Sprecher>  <Ort>  <Koordinaten>  <Geschl.>  <Geografische_Länge>
91  <Sprecher>  <Ort>  <Koordinaten>  <Geschl.>  <Geografische_Länge>
92  <Sprecher>  <Ort>  <Koordinaten>  <Geschl.>  <Geografische_Länge>
93  <Sprecher>  <Ort>  <Koordinaten>  <Geschl.>  <Geografische_Länge>
94  <Sprecher>  <Ort>  <Koordinaten>  <Geschl.>  <Geografische_Länge>
95  <Sprecher>  <Ort>  <Koordinaten>  <Geschl.>  <Geografische_Länge>
96  <Sprecher>  <Ort>  <Koordinaten>  <Geschl.>  <Geografische_Länge>
97  <Sprecher>  <Ort>  <Koordinaten>  <Geschl.>  <Geografische_Länge>
98  <Sprecher>  <Ort>  <Koordinaten>  <Geschl.>  <Geografische_Länge>
99  <Sprecher>  <Ort>  <Koordinaten>  <Geschl.>  <Geografische_Länge>
100 <Sprecher>  <Ort>  <Koordinaten>  <Geschl.>  <Geografische_Länge>
    
```

*AGD: Archive for Spoken German (<https://agd.ids-mannheim.de>)

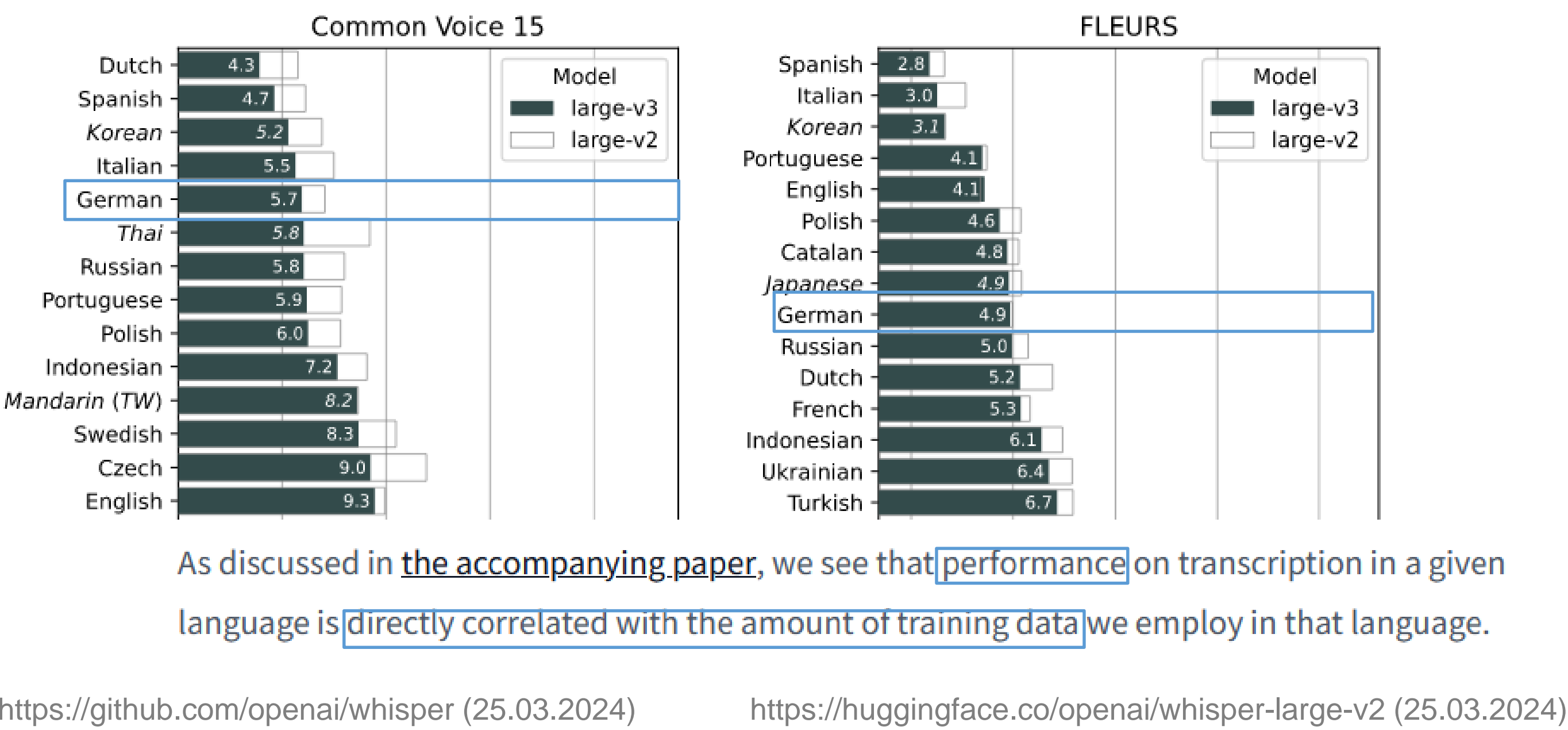
Problem: the transcription bottleneck



- Estimated need for transcription: 75.000h
 - e.g. 15 students for 10 years
 - + Overhead for management, quality control, technical supervision, documentation
 - Advanced dialect-competence(s) necessary – Alsatian, Low-German, Silesian, (data also contain Frisian, Sorbian, Dutch)
- Central, manual transcription in principle too expensive, not organisable
- Alternatives (cf. Brinckmann 2009)
 - Outsourcing
 - „Crowd“-Sourcing
 - (partial) automation



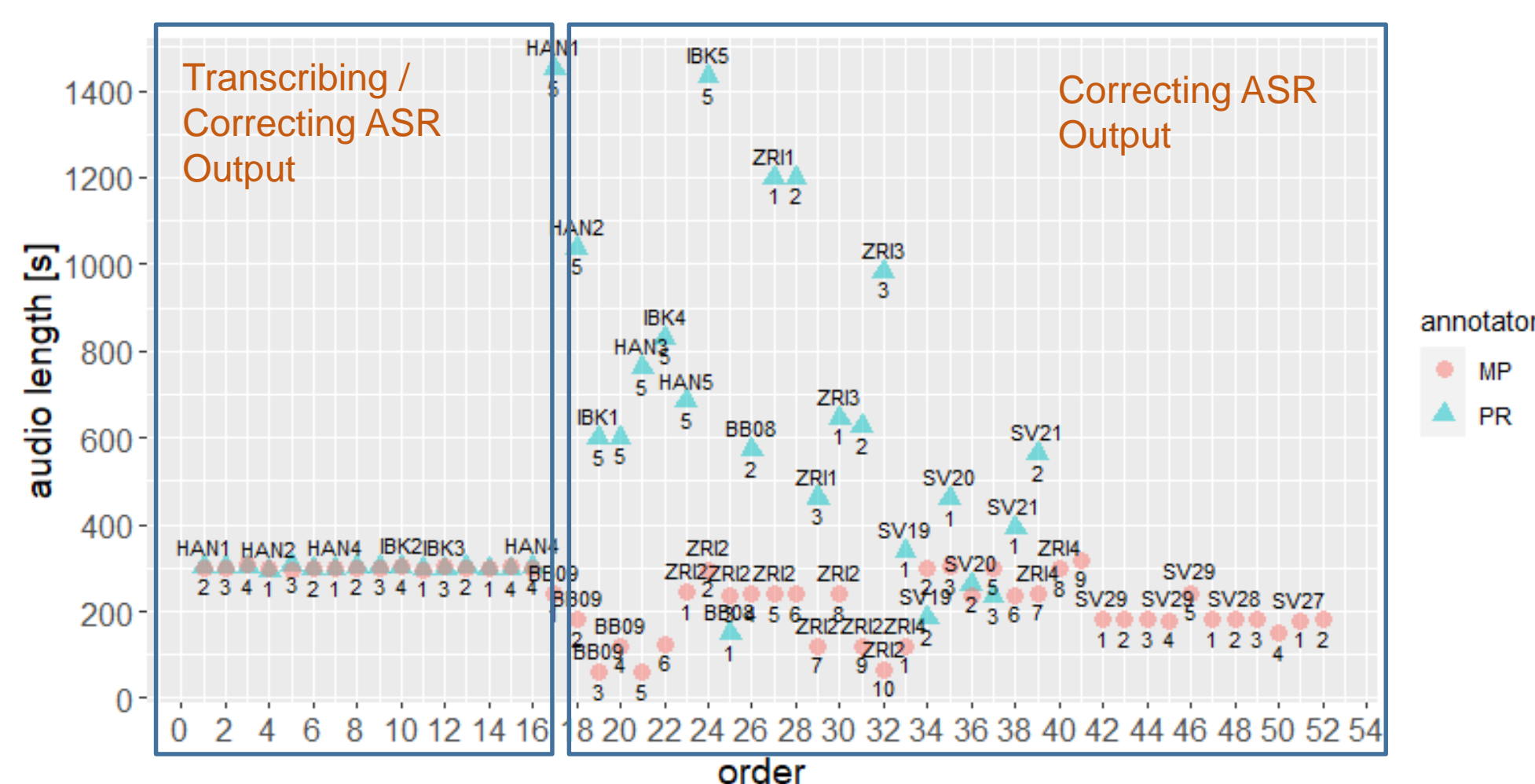
Word-Error-Rate (WER)



As discussed in the accompanying paper, we see that performance on transcription in a given language is directly correlated with the amount of training data we employ in that language.

<https://github.com/openai/whisper> (25.03.2024) <https://huggingface.co/openai/whisper-large-v2> (25.03.2024)

Findings



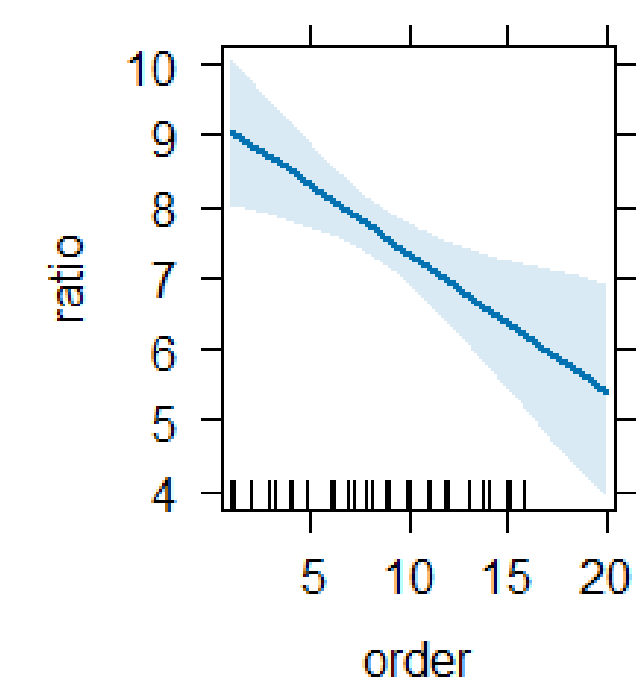
“Create orthographic transcripts”

- Alignment
- Mask tier

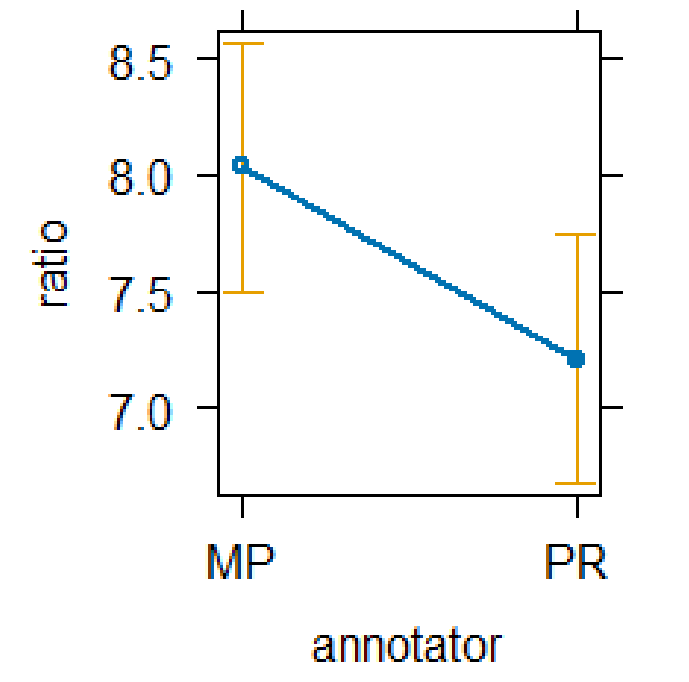
“Note the (working) time spent per stretch of recording”

Metric: ratio of $\frac{\text{audio length}}{\text{working time}}$

order effect plot



annotator effect plot

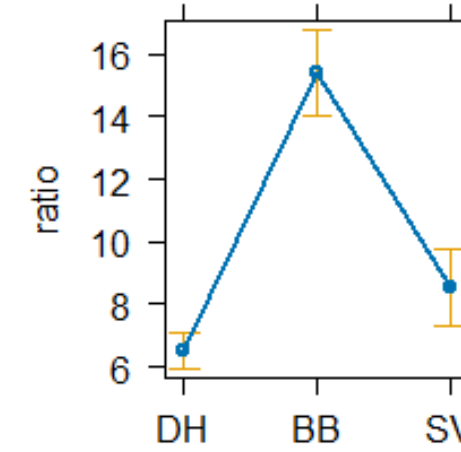


Factor	Estimate	Std. Error	z	p
(Intercept)	9.4537	0.4913	19.240	< 2e-16 ***
taskT	-0.1105	0.3705	-0.298	0.76780
order	-0.1923	0.0618	-3.115	0.0043 **
placeIBK	0.5344	0.5685	0.940	0.3555
annotatorPR	-0.8195	0.3692	-2.220	0.035 *

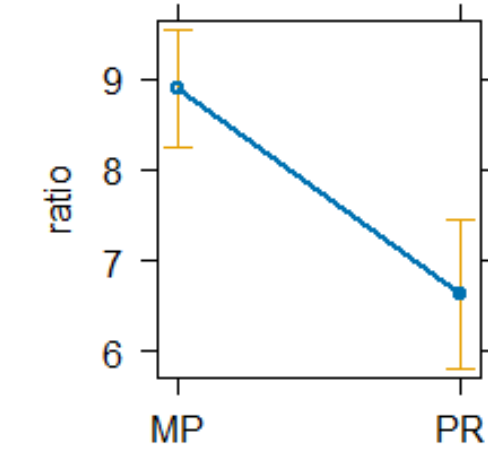
Adjusted R-squared = 0.33

Difficult to report a non-effect, but it seems that: Correcting and editing ASR-output takes about the same effort as creating transcriptions “manually” from scratch.

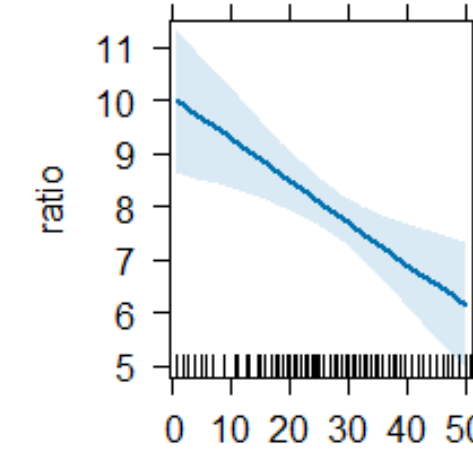
corpus effect plot



annotator effect plot



order effect plot



Factor	Estimate	Std. Error	z	p
(Intercept)	9.451	0.691	13.679	< 2e-16 ***
corpusBB	8.888	0.7327	12.131	< 2e-16 ***
corpusSV	2.0189	0.7552	2.637	0.0094 **
annotatorPR	-2.2656	0.5945	-3.811	0.0003 ***
order	-0.0787	0.0244	-3.232	0.0019 **
audio_length_in_s	-0.0004	0.00098	0.374	0.7094

Adjusted R-squared = 0.75

Research questions

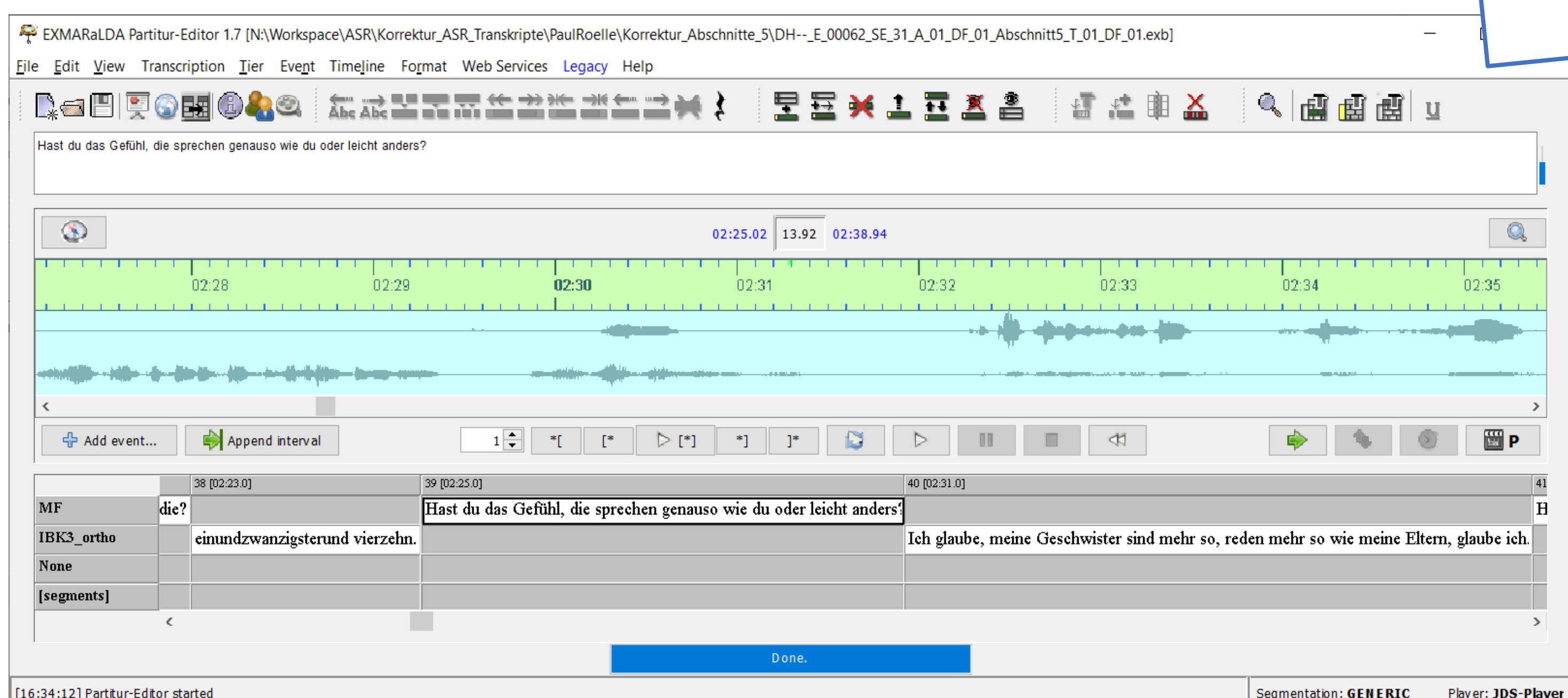
Huge improvements in ASR-quality in recent years (months)

- When is ASR-output good enough? (for the purpose of accessing oral corpora)
- If ASR-output needs to be corrected... is correction more efficient than transcribing from scratch?

We might now be at a tipping point... where correcting outperforms transcribing

Is ASR the solution to the problem?

Or do we need a different problem to the solution?



Acknowledgements

We would like to thank our two annotators Maja Peer and Paul Rölle for their annotation work and for insightful feedback on the data and the ASR performance. Many thanks to Sandra Hansen and Sascha Wolfer for their helpful advice on the statistical analysis.

Conclusions and Future Work

- Improve and optimize the solution:
 - Exploit prompting mechanism
 - Avoid some normalizations
 - Different post-processing
 - Adapt system to data (fine-tuning)

Adjust or redefine the problem:

- Analyze the imperfect output
 - Use additional ASR information, e.g. probability of word-detection
1. Generic query on ASR-transcribed corpus
 2. Refine manually the results