



**GHENT
UNIVERSITY**

TRANSKRIBUS IN FLANDERS/ BELGIUM

Dr. C.A. (Annemieke) Romein – 4 April 2019

FLANDERS & “EMDUTCH”

- Type of sources: political-legal texts
- Challenges: multiple languages (Dutch/French)
multiple hands (officials...)
- Tested on HTR: 15% CER
- Tested on HTR+: 7,5% CER (by Hr. Günter Hackl)

Details

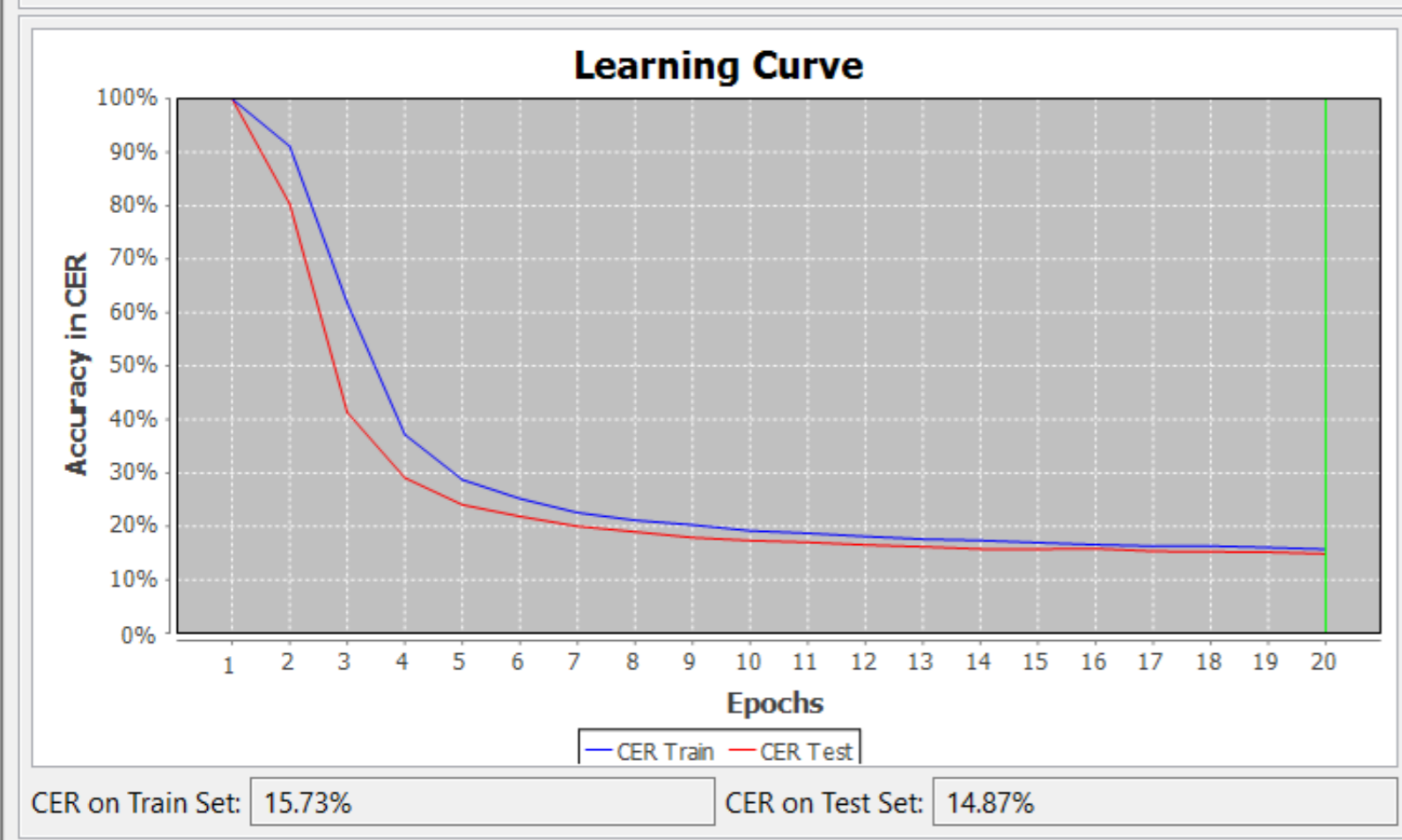
Name: Language:

Description:

Parameters:

Nr. of Epochs	20
Learning Rate	1e-3
Noise	both
Train Size per Epoch	10000

Nr. of Words: Nr. of Lines:



Details

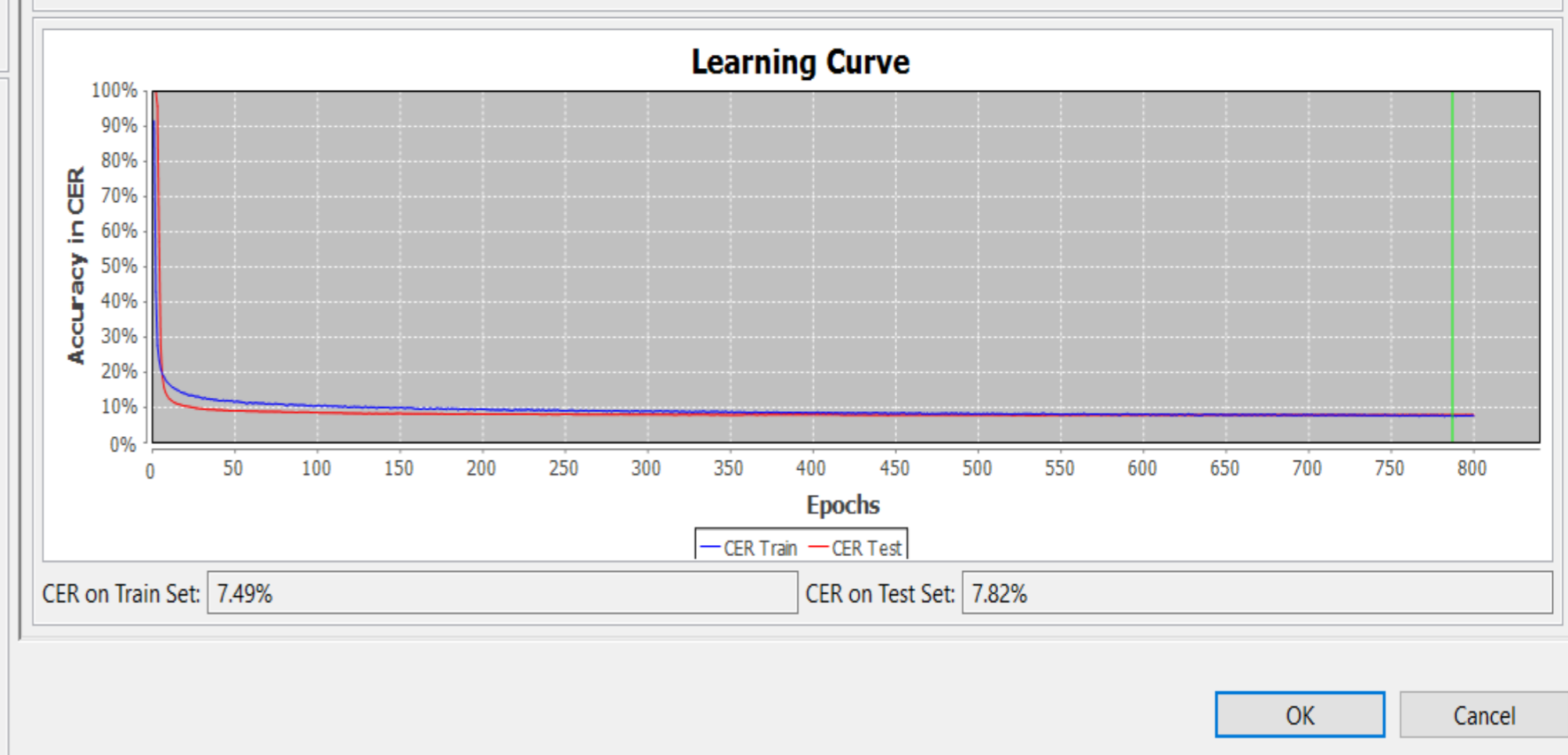
Name: Language:

Description:

Parameters:

Nr. of Epochs	800
---------------	-----

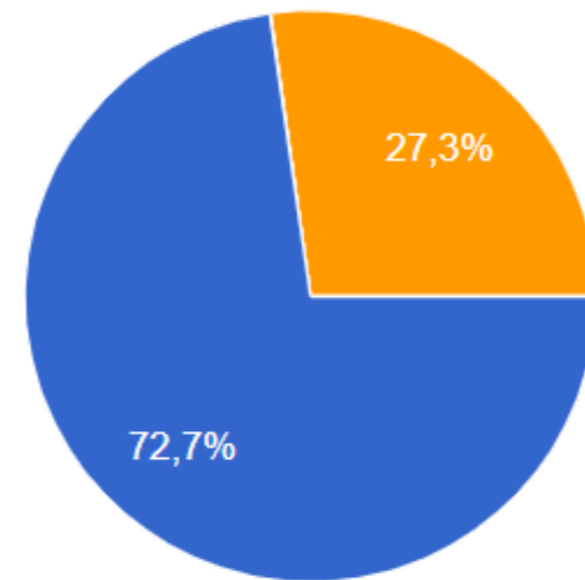
Nr. of Words: Nr. of Lines:



A WARM UP FOR THE DISCUSSION...

Are you prepared to share your data (transcripts)

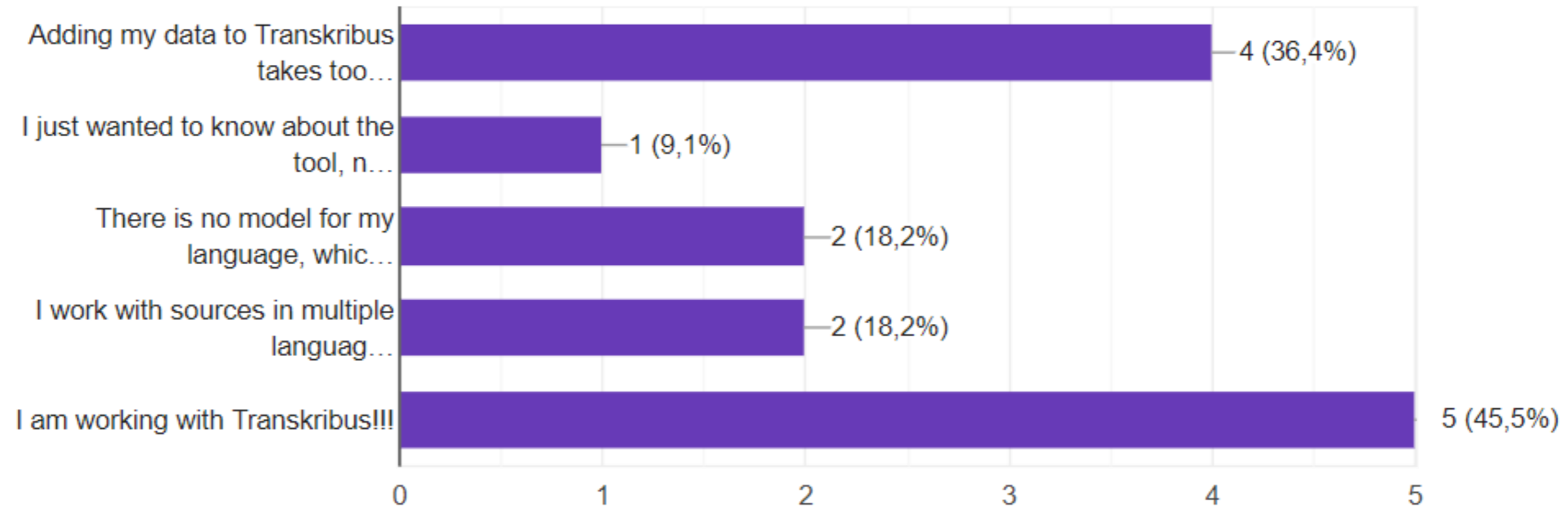
11 antwoorden



- Yes
- No
- May be: only a computer would be allowed to 'learn' and create a model with my data (not a human person 'prying').

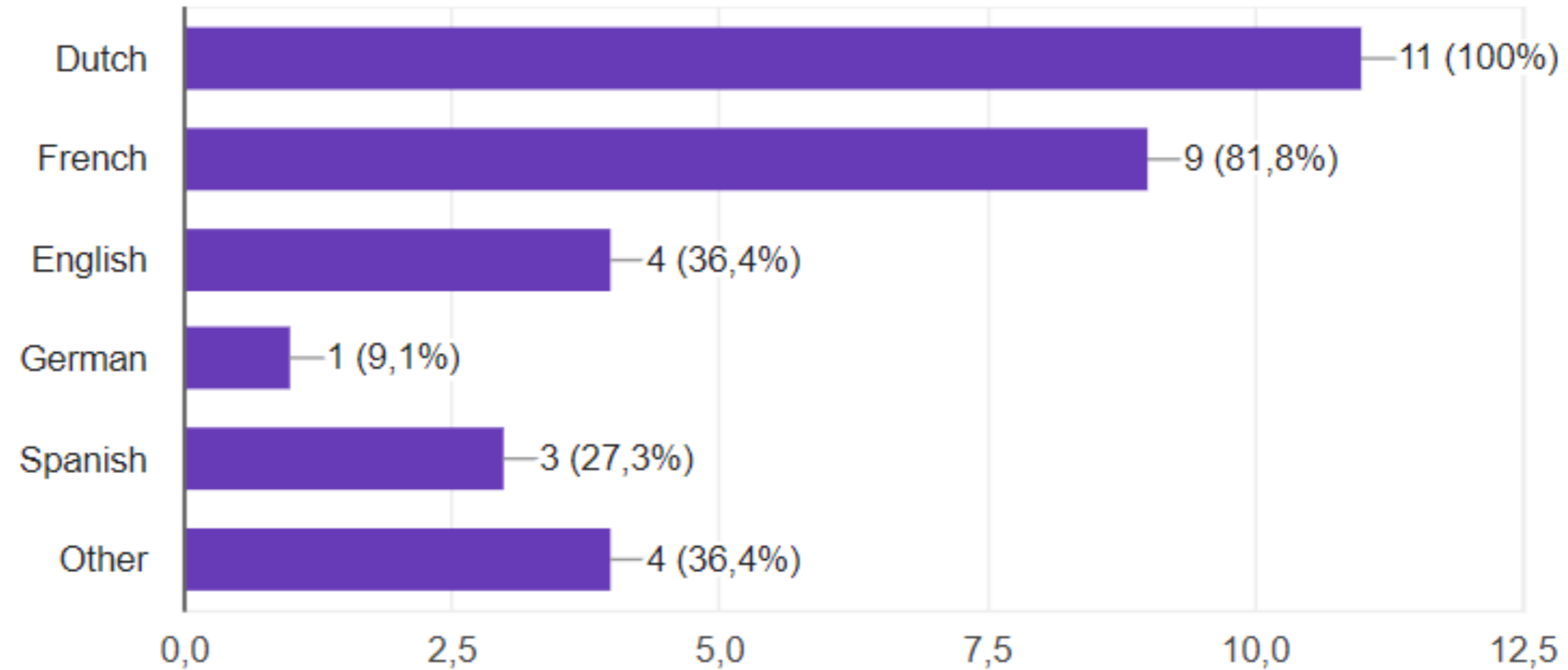
What is the reason you are not working with Transkribus?

11 antwoorden



What language are your sources in? (multiple options)

11 antwoorden



WILL YOU BE USING TRANSKRIBUS IN THE
NEAR FUTURE? IF YES: WHAT DO YOU THINK
IS USEFUL ABOUT THE TOOL; IF NO, WHAT IS
HOLDING YOU BACK?

- My sources are typewritten and can be ocr'd but some of them are hard to “read”, which is why Transkribus could be useful.
- Yes, but mainly the OCR options because I work with newspapers.
- Useful: seems to work very well with printed sources.
- Yes, but it depends on the availability of a proper (basic) model! (If I have to train a model, I should get a discount!)
- Yes. As an archivist we look forward to the automatic enrichment of data.

HOPES AND MISSED CHANCES...?

- We zouden graag Transkribus willen gebruiken voor de 18de eeuwse Dietse akten uit de schepenbank evenals voor de transcriptie van een 18de eeuwse kroniek.
- Ik heb een wetenschappelijk project ingediend bij de Belgische federale overheid over de ontsluiting van het archief van de Raad van Brabant (15de-18de eeuw). <Currently under review>
- I do intend to use Transkribus; I even mentioned it in a FWO-project proposal.

- I discovered Transkribus too late: my 250.000 word transcription is already finished. Putting it into Transkribus now seems loss of time.

TEXTUAL GENRE & ACCURACY (PREDICTABILITY)

- (1) I would be interested to know out how important the factor 'textual genre' is for the accuracy with which a model can predict words.
- (2) We moeten toewerken naar één generiek model. Aandachtspunten zijn o.m. verschillende periodes en waar de grens trekken (of één model voor alle periodes?) en meerdere talen binnen één type document/model. In mijn geval Nederlands-Latijn-Frans.

Dr. C.A. Romein

Postdoctoral Researcher Early Modern Political-Institutional History

HISTORY

E Annemieke.Romein@ugent.be