CLARIN Resources for Classical Latin and Historical German

Brian MacWhinney
Professor of Psychology, Carnegie Mellon University
macw@cmu.edu

CLARIN Annual Conference 2016, Aix-en-Provence
26–28 October 2016

CLARIN

Latin and Early Modern German

- LATIN
  - Latin is a closed corpus of 454 texts
  - Latin learning has not yet entered the digital age
  - Alpheios methods died with the Firefox plugin
  - Major challenge: token-based alignment in JSON
  - After this, making web functions will be
- Early Modern German (1420 – 1750)
  - This is not yet a fully closed corpus
  - Because there is still major OCR work to be done
  - Lexical normalization is a major and interesting task
  - Sentence segmentation is a major and interesting task

Not quite your usual kind of resource. Gra.fo and the documentation of Oral Archives in CLARIN

SILVIA CALAMAI & FRANCESCA FRONTINI
Università degli Studi di Siena; Université de Montpellier 3 – Paul-Valéry
silvia.calamai@unisi.it; francesca.frontini@univ-montp3.fr

CLARIN Annual Conference 2016, Aix-en-Provence

26–28 October 2016

CLARIN CLARIN-IT

Gra.fo & CLARIN

- We present some reflections on the documentation of **Oral History archives** within CLARIN with a focus on their **accessibility** through the CLARIN Virtual Language Observatory.
- The case study → «**Grammo.foni. Le soffitte della voce» (Gra.fo) project**, a collection of digitized and catalogued oral Tuscan archives.

grafo.sns.it

- Come and see our poster if you are interested in:
  - **Oral history, oral archives, oral resources**
  - **Metadata curation**

Researcher Hands-On Training in the Digital Humanities: An Austrian Case Study

TANJA WISSIK & CLAUDIA RESCH

Austrian Academy of Sciences, Austrian Centre for Digital Humanities

tanja.wissik@oeaw.ac.at claudia.resch@oeaw.ac.at

CLARIN Annual Conference 2016, Aix-en-Provence

26–28 October 2016

CLARIN

Morning lectures

&

Exchange of experiances regarding training – visit our poster in the poster session 1 16:30 - 17:30!

Hands-on afternoon

AHA: Anagram Hashing Application

Martin Reynaert

Center for Language and Speech Technology - Radboud University Nijmegen
TiCC - Tilburg University

CLARIN Annual Conference Aix-en-Provence 27 October 2016

Faculty of Arts
Centre for Language and Speech Technology
Radboud University Nijmegen


AHA: Anagram Hashing Application

- We present AHA, the Anagram Hashing Application, a new web application and service.
- AHA allows researchers to effortlessly analyse the lexical variation present in their Gold Standard data and to publish the results.
- AHA gives statistics on the lexical variation present in the Gold Standard.
- AHA gives the list of character confusions and their frequencies.

AHA is a subsystem in PICCL

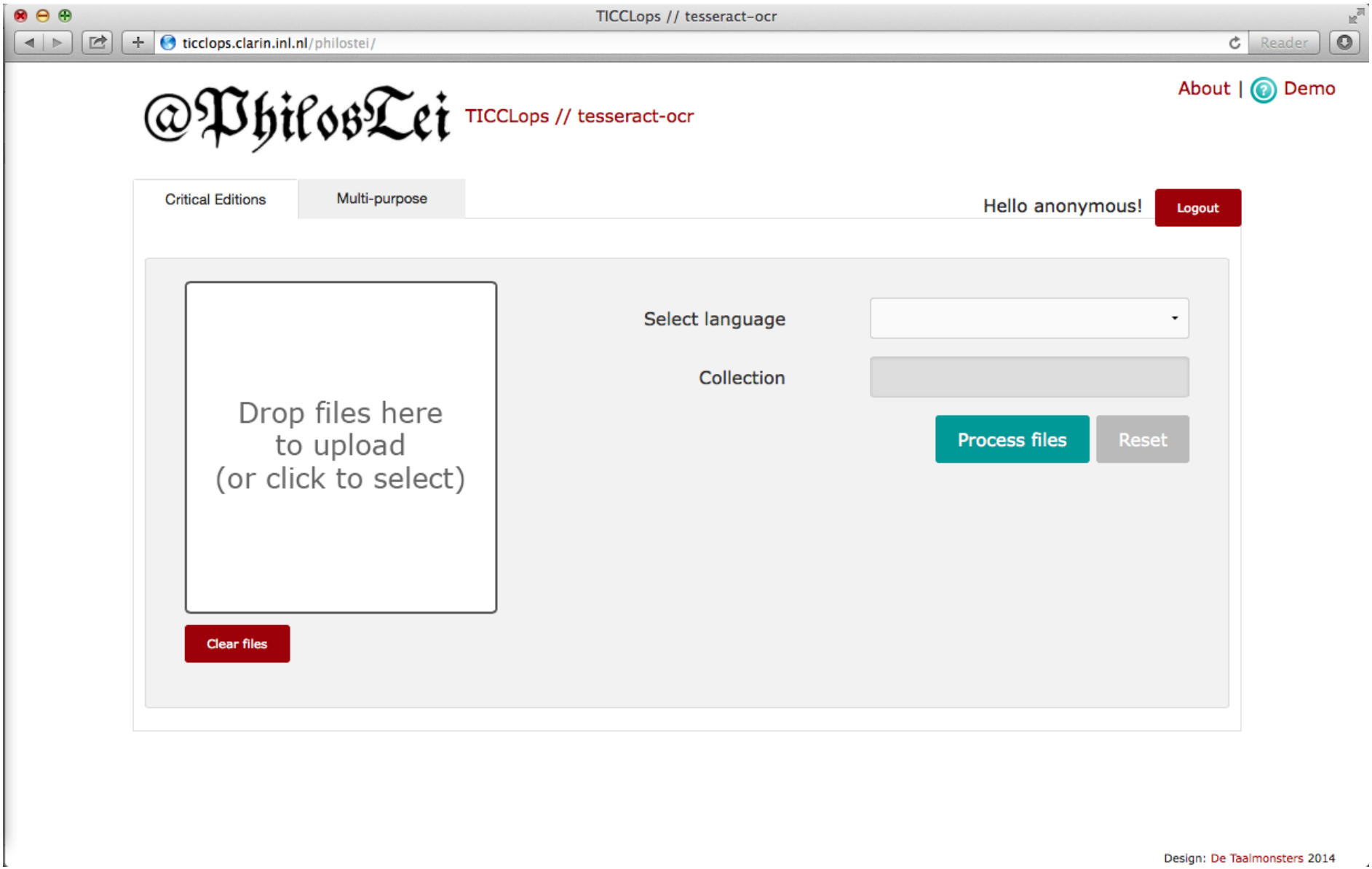
- We present a new corpus building tool called PICCL. It constitutes a complete workflow for corpus building.
- PICCL is to be the integrated result of recent developments in the CLARIN-NL project @PhilosTEI, which ended November 2014, and further work in NWO ‘Groot’ project Nederlab, which continues till end 2018 and in CLARIAH, which will run till 2020.
- PICCL wants to move beyond demonstrator status and be an actual production system.

PICCL: An integrated pipeline

The integrated PICCL pipeline offers:

- a comprehensive range of conversion facilities for legacy electronic text formats.
- Optical Character Recognition for text images: Tesseract.
- automatic text correction and normalization: TICCL.
- linguistic annotation with Frog: tokenisation, lemmata, POS, NER.
- and indexing for corpus exploration and exploitation environment WhiteLab.

PhilosTEI screenshot



ENJOY!!

Thanks for your attention!

`http://philostei.clarin.inl.nl/`

AHA: Anagram Hashing Application

Martin Reynaert

Center for Language and Speech Technology - Radboud University Nijmegen
TiCC - Tilburg University

CLARIN Annual Conference Aix-en-Provence 27 October 2016