WG1 Meeting
21th of April 2022, our last one

Who is attending:

- (partially) Diana
- Serge Heiden
- Borja Navarro
- Lou Burnard
- Martina Scholger
- Carolin Odebrecht
- Cvetana Krtev
- Giulia Grisot
- Michal Kren
- Michael Preminger

Agenda

1.Welcome (done)
Thanks to WG1 and everybody who was involved in creating ELTeC! WE did a great job and created a ressource that is used and will put a benchmark to others. :-)

ELTeC
- https://distantreading.github.io/
- ELTeC https://distantreading.github.io/ELTeC/index.html
- Data Repositories: https://github.com/COST-ELTeC
- Zenodo Releases https://zenodo.org/communities/eltec
- Final Corpus Release Mai 2022

2. GAMS
For fostering the sustainability and visibility of ELTeC, we will use GAMS. It is an IT-Service for storage, visualisation of TEI Collections.
https://gams.uni-graz.at/ (Martina presenting)

Which Releases/Levels/Core/Extension
1. only current the current version
2. which Levels
    a. for display we use level0 or level 1 (if available), links to zenodo doi for every novel
    b. Link to TXM would be great (like we do have a link to Voyant) - Future plan
    c. And link to CWB :) as future…
    d. Or remove voyant tools? No (or maybe yes until we have more options like TXM and CWB (Diana))
    e. Another nice/widely used tool is Antconc
3. Core - Yes
4. Extension? see below

Suggestions/Questions:
- Combining Languages for searching in ELTeC would be good to have
- Translation for GAMS (Serge is helping! Diana can also help for Portuguese)

<mark>Publication of ELTeC in GAMS by the end of June/ beginning of July</mark>

3. TXM- Interface

- Gitlab repository: https://gitlab.huma-num.fr/sheiden/ELTeC-fra
- Importing TEI sources into TXM
  - Specifications import ELTeC-FRA into TXM document (level 1)
    -> to help import ELTeC corpora for any language
    - Document Unit
    - Metadata
    - Document Structure
    - Textual planes
    - Primary language
    - Lexical units
    - Text Edition
    - Editorial Unit
    - Technical specification of Text Rendering
    - Localization References
- Using TXM ELTeC corpora
  - ELTEC-FRA-L1-V1.txm (first ELTeC TXM corpus test)
    - tokenization & lemmatisation done by TXM
    - downloadable for TXM for desktop (Winows, Mac & Linux)
    - Hosted in future txm-eltec TXM portal hosted in huma-num.fr
      -> need to connect this to CLS INFRA project
  - ELTEC-FRA-L2
    - prototype corpus being constructed (Sascha Diwersy + Lyon TXM team)

4. Extension and Core

How do we distinguish between core an extension? How do we make it clear to users that the data sets are two different but related things?

Core Collection
        Design and criteria behind the creation,
        (being a prose text of a certain length and certain time etc.)
Extension collection includes data
        Without meeting the design criteria (different reasons)
        Not conferment with the Design/schema
        Limited Progress (limited text amount)

Lou will provide examples how to structure those data sets and send it via wg1 mailing list

5. AOB
Thanks to WG1! :)