

PeregrXML:

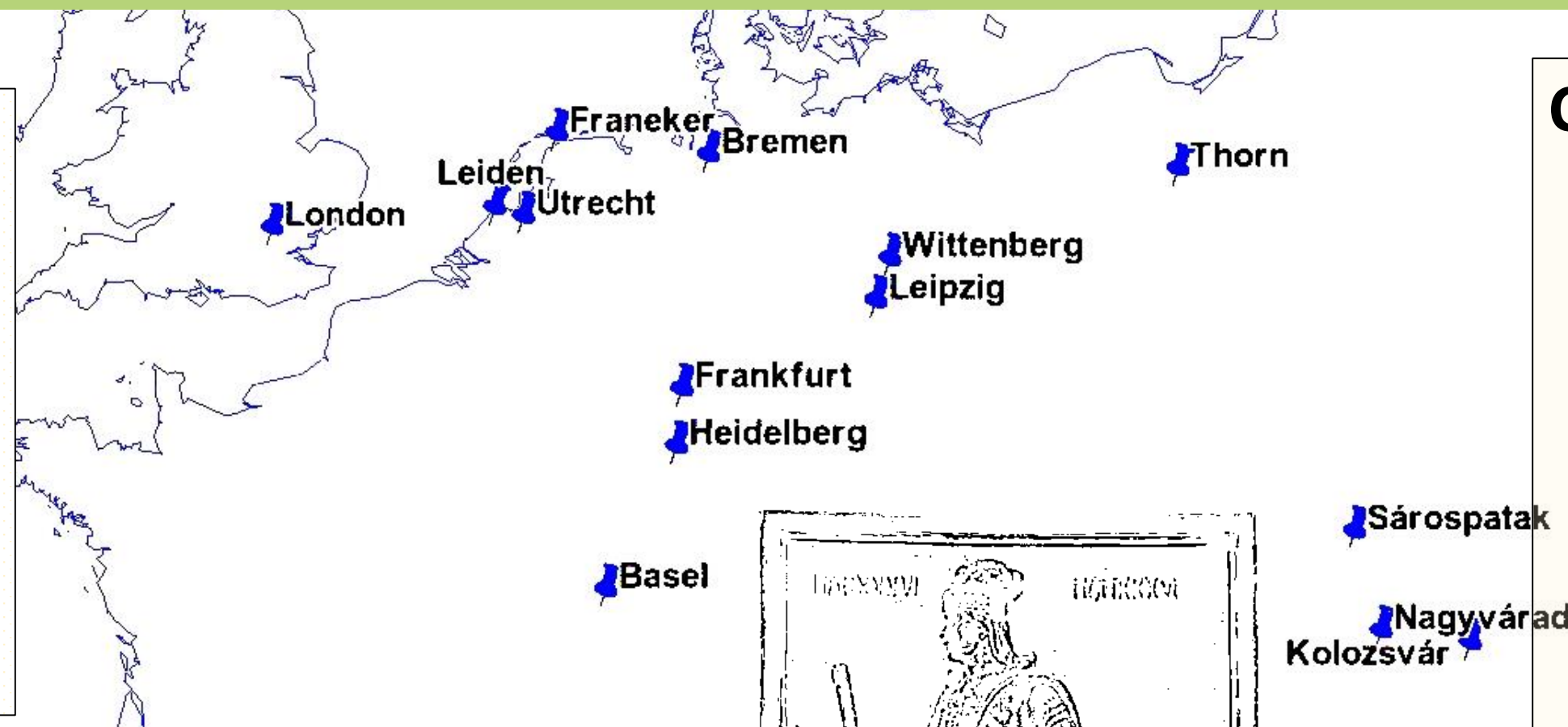
Building and annotating a 17th century corpus of Medieval Christian Hebrew

Tamás Biró^{1,2} joint work with: Andrea Götz³, Szandra Juhász³, Kornélia Koltai^{1,3} and József Zsengellér³

¹ ELTE Eötvös Loránd University, ² Jewish Theological Seminary – U. of Jewish Studies (Bp.), ³ Károli Gáspár University of the Reformed Church



Background: Our project focuses on **gratulatory poems** written in Hebrew by Hungarian protestant students studying theology at Dutch and German universities (“peregrines”) in the seventeenth century. These poems – together with similar poems written in Latin, Greek, Syriac and other languages – were recited at public events, such as theses defenses, and published in print subsequently.

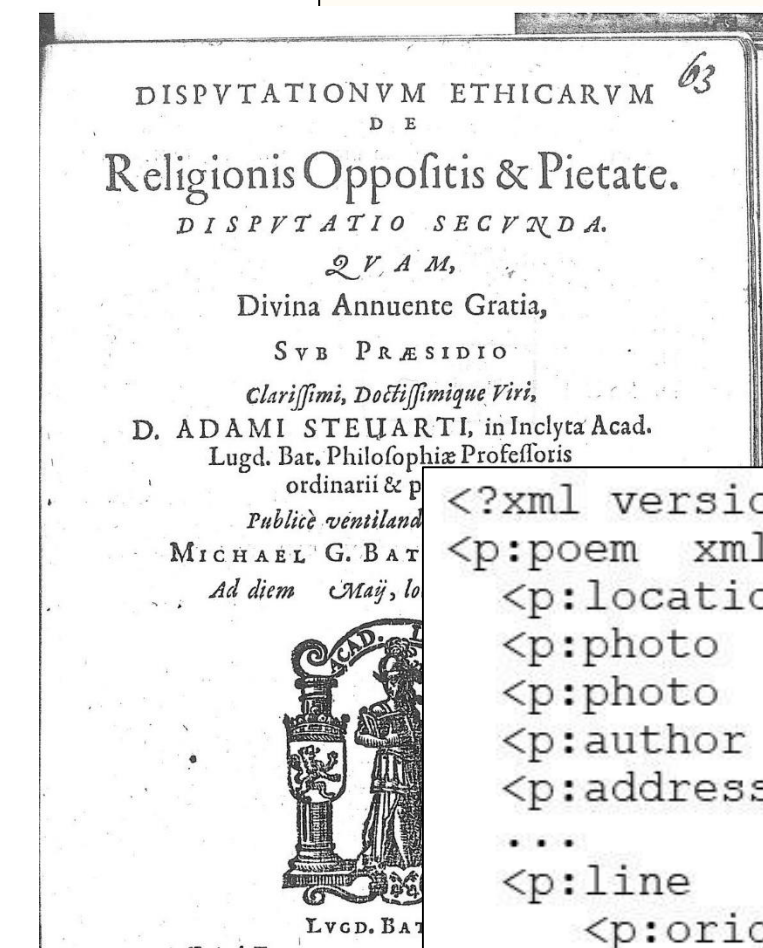
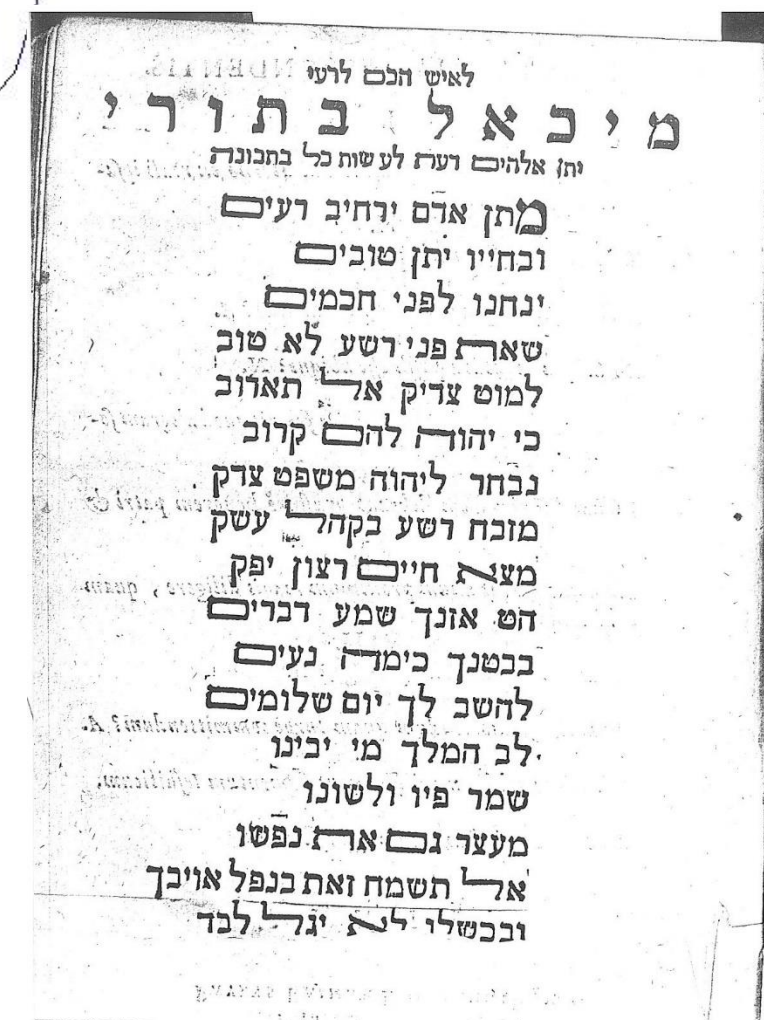


Challenges:

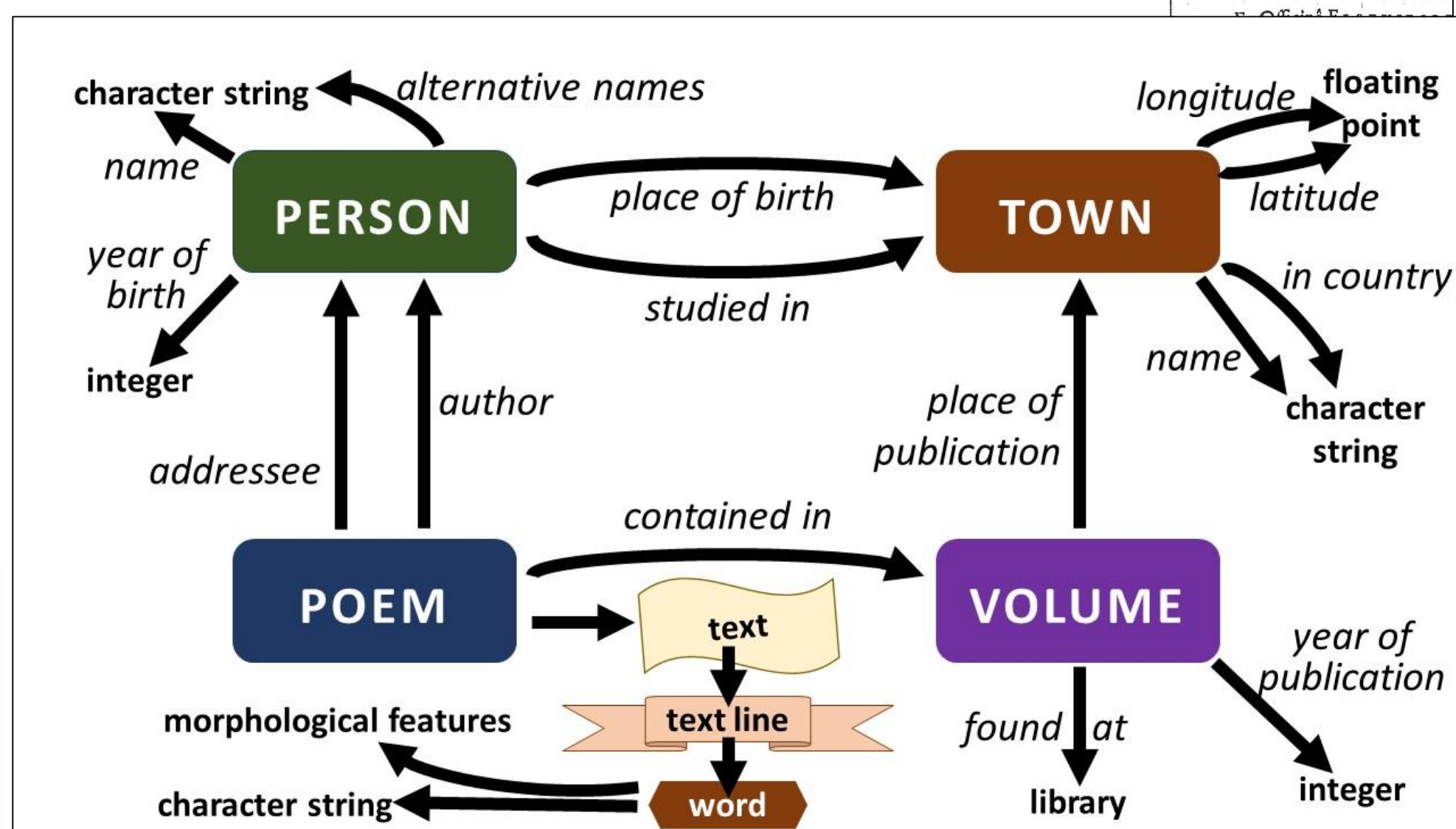
1. Quality of the photocopies. Typos in the original texts.
2. Imperfect language skills of the authors.
3. Various encodings: Hebrew script with vocalization, Hebrew script without vocalization, Latin script.
4. Hebrew without vocalization: several readings.
5. Uncertainties and multiple readings on several levels.
6. Normalization: correct for imperfect language skills of the authors? Depends on research questions.
7. A broad array of (possible) research questions:
 - a. Language skills of the authors.
 - b. (Intended) language of the poems.
 - c. Poetic structures, literary aspects.
 - d. Cultural, historical, theological aspects.

Phases of the project:

1. Collecting the texts from various libraries.
2. Transcription, normalization, translation.
3. Answering research questions, such as
 - Does the grammatical knowledge of students at different universities display significant differences?
 - How are certain linguistic features used, compared to Classical Biblical Hebrew, Late Biblical Hebrew and Post-Biblical Hebrew? To grammars in that time?
 - Which Biblical books serve as sources? (Linguistic patterns copied, intertextuality, paraphrases...)
 - Are theological concepts, historical events or personal episodes being referred to?
 - What kinds of social networks are formed by the authors and the addresses?
 - Are there cases of L1 or L2 language transfer in these texts produced in a language foreign to the authors?



```
<?xml version="1.0" encoding="UTF-8"?>
<p:poem xmlns:p="peregr.kre.hu">
  <p:location book_id="Leiden5_RMK_III_1840" date="1653" />
  <p:photo src="img5.jpg" />
  <p:photo src="img7.jpg" />
  <p:author id="dkp" name="Diószegi Kalmár Pál" />
  <p:addressee id="bm" name="Báthori Mihály" />
  ...
  <p:line id="13">
    <p:original >הַט אֲזוֹנֵךְ שָׁמַע דְּבָרַיִם</p:original>
    <p:normalized>הַט אֲזוֹנֵךְ שָׁמַע דְּבָרַיִם</p:normalized>
    <p:transl_hu >Fordítsd füledet, hallgasd a szavakat</p:transl_hu>
  </p:line>
  <p:line id="14">
    <p:original >בִּוְטוֹךְ כִּיחָה נָעִימִ</p:original>
    <p:normalized>בִּוְטוֹךְ כִּיחָה נָעִימִ</p:normalized>
    <p:transl_hu >A bensődben, kellemes fiastyúk</p:transl_hu>
    <p:normalized>בִּוְטוֹךְ כִּי חָה נָעִימִ</p:normalized>
    <p:transl_hu >A bensődben mily kedves</p:transl_hu>
  </p:line>
  <p:line id="15">
    <p:original >לֵךְ יוֹם שְׁלוֹחִים</p:original>
    <p:normalized>לֵךְ יוֹם שְׁלוֹחִים</p:normalized>
    <p:transl_hu >Hogy visszahozza számodra a jutalom napját.</p:transl_hu>
  </p:line>
  ...
</p:poem>
```



Type of intended queries:

```
SELECT morphological_features FROM poems
WHERE poem.author.studied_in.name == Leiden
```

Conclusions? Not yet. Still work in progress.
However, any suggestion very much appreciated.