Data Science for Hard Core Humanists.

Opportunities and Challenges in Computational Assyriology

Wed March 4 2020

**Slide 2**

Compass: examples of computational methods for Assyriology – the study of cuneiform.

**Slide 3**

Revolution in Information Technology. South of Iraq, city of Uruk, 5,000 years ago: writing:

**Slide 4**

Writing initially for commercial and admin purposes – still common in IT developments!

Over time: other purposes: royal inscriptions, songs and myths, divination, astronomy, mathematics, medicine.

Sumerian: linguistic isolate, Akkadian: Semitic language – related to Arabic and Hebrew

**Slide 5**

Four representations of the same simple sentence in Sumerian: cuneiform, transliteration, lemmatization, translation.

**Slide 6**

Collecting digital data: since 1996.

**Slide 7**

Institutional developments and open source projects

**Slide 8**

Potential of Data Science and Assyriology meeting:

* Scaling up

Old Babylonian scribal school:

* Lexical lists
* Copying Sumerian literary texts

2004: compare *bird* vocabulary by hand

Now: use the entire corpus of literary and lexical texts

**Slide 9**

* Social Network Analysis

SNA Already done on *private* data

Majority of texts administrative

**Slide 10**

Topic model: primarily for exploratory purposes. For teaching?

Vocabulary perspective: probability distribution of words over topics

Document perspective: probability distribution of topics over documents

**Slide 11**

Word embeddings in Fasttext

Word vectors, built from entire Sumerian corpus – about 3 million words

**Slide 12**

Will my colleagues be interested in using these tools?

How does it fit into their work flow?

**Slide 13**

Multivalency of cuneiform signs

**Slide 14**

Mash up of global sign list (all sign values) and BDTNS (very large group of administrative texts)

Eyes lighting up: possibility to look for SU followed by HI

More data wrangling than data signs

User Interface made with ipywidgets

**Slide 15**

Unpredictable behavior when modules interact

incomprehensible set-up commands

**Slide 16**

Changing attitude towards open data