A Time to Every Purpose under Heaven:

Extracting Dates from Hebrew and Aramaic Texts

Jelte van Boheemen, José de Kruif, Ortal-Paz Saar, Sheean Spoel

Authors' affiliation: Utrecht University, Netherlands

Presentation category: Short paper (10 minutes)

Hebrew and Aramaic historic texts, such as funerary inscriptions from 500 BCE onward, rabbinic and responsa literature, or Memorbücher, are rich in temporal information. Many such corpora are digitized, yet there is no efficient system to automatically extract the temporal information (explicit or relative) from them.[1] While there are a few date parsers that support Hebrew,[2] they function with modern texts, using standard notations and Arabic numerals. However, they cannot be used on historic texts. In these latter, dates are invariably expressed through Hebrew letters, not numerals. For example, the number 14 is written as two letters, *yod* and *dalet*, יד. When taken as a word, this two-letter sequence may also be read as *yad*, "hand". Additionally, the notation standards vary widely, for instance: "since the Temple destruction", "in the reign of King Seleucus", or "the third year of the seven-year cycle".[3] This variation makes it even more complicated to recognize what date is mentioned in the text. The present paper will describe the development of a novel algorithm suited to automatically extract dates from historic Hebrew and Aramaic texts.

Aiming to produce a Python module and application to automatically mark, extract and standardize these dates, the following method is employed:

> 1. In collaboration with historians from Judaic Studies, a formal representation of Hebrew and Aramaic dates has been created, including specific patterns of temporal representation. These patterns take into consideration that historic texts, especially inscriptions, are often damaged, which results in incomplete temporal sequences.[4]

---

[1] For temporal tagging in other fields see, e.g., Strötgen and Gertz 2012.

[2] See, e.g., https://dateparser.readthedocs.io, https://github.com/ishirav/date-detector. Limited support for Aramaic is provided by HeidelTime (https://github.com/HeidelTime/heideltime); however, it offers no support for non-Gregorian dates.

[3] For a basic introduction to the Hebrew numerical system, see, e.g., Chrisomalis 2010: 156-160.

[4] E.g., a surviving inscription may read "in the [....] seven hu[........]fty three since the des[..........]", which should be restored as "in the [year] seven hu[ndred and fi]fty three since the des[truction of the Temple]"; and then understood to mean year 823 of the Gregorian calendar.

2. Creating conversion scripts for converting the dates found to standardized forms, mapping them to Gregorian and GODOT dates.[5]

3. Creating a text parser to search texts for these patterns.

4. Verifying the parser by comparing its results with manually annotated text corpora, e.g. the 35,000 Hebrew funerary inscriptions of the Epidat database.[6]

5. Making the algorithm freely available on Github.

Having such dates in standardized format allows searching for historical, religious, linguistic and even astronomically-related data efficiently, including by those with no knowledge of Hebrew or Aramaic. Furthermore, the tool is expandable for additional languages, particularly Arabic texts, which contain dates in similar formats.[7] Additionally, the tool will allow crossing dates of events mentioned in different, unrelated corpora, including those in languages other than Hebrew and Aramaic, through systems like PeriodO.[8] The tool can also enrich existing corpora with useful information about their dating, and may thus provide insight into patterns (e.g. migration, cultural development) on a historical timeline. In other instances, seemingly unconnected events may be found to be historically linked. These elements boil down to enabling researchers to discover patterns and connections previously undetected, in other words: show that there is a time to every purpose.

---

[5] See https://godot.date/home.

[6] See http://www.steinheim-institut.de/cgi-bin/epidat.

[7] For some interesting past work on historical calendrical conversions, see Dershowitz and Reingold 1990; Reingold, Dershowitz, and Clamen 1993.

[8] See https://perio.do/en/.

Bibliography

Chrisomalis S. 2010. *Numerical Notation: A Comparative History* (Cambridge and New York: Cambridge University Press).

Dershowitz N. and Reingold E.M. 1990. "Calendrical Calculations", *Software: Practice and Experience* 20.9, pp. 899-928.

Strötgen J. and Gertz M. 2012. "Temporal Tagging on Different Domains: Challenges, Strategies, and Gold Standards", *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC 2012)*, pp. 3746-3753.

Reingold E.M., Dershowitz N., and Clamen S.M. 1993. "Calendrical calculations, II: Three historical calendars", *Software: Practice and Experience*, 23.4, pp. 383-404.