

# The Arabic Papyrology Database

Johannes Thomann

University of Zurich, Switzerland

## Abstract

There exist about 150,000 premodern Arabic documents on papyrus and paper, of which about 2,500 have been edited. Another 10,000 unpublished documents are described or mentioned in papyrological publications. It is the aim of the Arabic Papyrology Database (APD) to give access to published texts and descriptions. For the APD, an entirely new approach of organizing Arabic text was developed. The APD presents texts in five levels that account for the peculiarities of the Arabic script system and the scribal practices. The first level provides a faithful diplomatic edition of the text as found in the document. The following levels document four steps of editorial interventions. On the second level, lines of characters are broken into single words. On the third level, lacking diacritical dots are supplied. On the fourth level, Arabic vowel signs are added, providing a full phonological representation. On the fifth level, a scientific Latin transliteration is given. Each element of the fifth level is connected to a lexicon and a list of grammatical forms. All levels contain variant readings and editorial remarks. At present, the APD contains 1,806 full-text documents and is freely accessible (<http://www.naher-osten.lmu.de/apd>).

## Correspondence:

Johannes Thomann,  
Institute of Asian and  
Oriental Research,  
Wiesenstrasse 9, CH-8008  
Zurich, Switzerland.  
E-mail: johannes.  
thomann@aoi.uzh.ch

## 1 Peculiarities of the Arabic writing system

The Arabic writing system in general and the writing conventions in papyri in particular require special display formats. As in most Afro-Asiatic writing systems, short vowels are not represented by letters, but there exist vocalization marks (*ḥarakāt*), written above or below the letters (Gacek, 2009: 288). Further, one-letter words and the article are written together with the following word, and the pronominal suffixes are attached to the preceding word. In transliteration, these words are separated by a hyphen. Finally, fifteen letters (*al-ḥurūf al-muʿjama*) are distinguished by diacritical pointing from letters with the same basic form (Gacek, 2009: 144). However, these dots are only occasionally found in early Arabic documents (Kaplony 2008). Most modern editions of early and classical Arabic texts do not account for these peculiarities of the Arabic writing system and normalize the texts according to

modern orthographical rules without using vowel-signs. While this established practice may be appropriate for literary texts, a more elaborate procedure is needed for documentary texts.

## 2 The data

There are two main groups of premodern Arabic documents: Inscriptions and papyri in the broader sense. The second group consists of about 150,000 documents written on different materials such as papyrus, parchment, or paper during the period from the 7th to the 15th centuries, of which about 2,500 have been edited (Sijpesteijn 2005). Another 10,000 unpublished documents are described or mentioned in papyrological publications. All these documents provide information on almost every aspect of Islamic history. Despite their importance, they still do not receive a high level of attention in historical research (Sijpesteijn 2009: 456).

### 3 Methodology

Two leading ideas founded the Arabic Papyrology Database (APD, accessed 1 October 2014: <http://www.naher-osten.lmu.de/apd>): on the one hand, a research tool which makes metadata and full texts of all edited documents easily accessible should be provided, and, on the other hand, the limitations of printed editions could be overcome by this digital edition. As already mentioned, modern editions present the texts in a normalized form according to modern Arabic orthography. Some high-quality editions describe the diacritical pointing and vocalization marks in the critical apparatus, and a limited number of editions provide word indices in transliteration.

For the APD, an entirely new approach of organizing Arabic text was developed. Instead of the single text-level approach used in print and in other database projects, the APD presents texts in five levels. On the first level, only diacritical dots found in the document are written and all observations of gaps, deleted texts, and redundant words are indicated by sigla and brackets. On the second level, these marks are removed and the text is broken into single words. On the third level, missing diacritical dots are added, as it is common in text editions. On the fourth level, vowel marks are added, providing a full phonological representation. On the fifth level, the text is given in scientific Latin transliteration; words written together in Arabic are now separated. Further, each element of the fifth level is connected to a lexicon and a list of grammatical forms. The levels of text are hierarchically organized, and variant readings and remarks are attached to their appropriate level. Search is not only possible along each level, but searches across levels for the corresponding or neighbouring word in another level can be carried out as well. Beyond these layers of text, some steps have been made to include a treebank of syntactical structures in the APD, based on the identified word categories and grammatical forms. Hopefully, its realization will become part of a future project. At present, pairs of adjacent

word categories with particular grammatical forms can be found in the texts.

### 4 State of the project

The APD was initiated by Andreas Kaplony and Johannes Thomann at Zurich University in 2004. Today it is a joint project of the Universities of Zurich, Munich (LMU), and Vienna. The APD was online and freely accessible from its beginning. A PhD project by one of its collaborators was based on the APD (Grob, 2010), and another research project would have been impossible without the advanced search capabilities across text levels (Kaplony, 2008). At present, 1,806 full text documents are available in the APD, and by 2016 the remaining edited documents will be entered into the APD. There are mutual references in the APD and the Trismegistos database, and soon, texts of the APD will be imported by the Papyrus Navigator, made possible by the XML export engine of the APD in EpiDoc format.

### References

- Gacek, A. (2009). *Arabic Manuscripts: A Vademecum for Readers*. Leiden: Brill.
- Grob, E.M. (2010). *Documentary Arabic Private and Business Letters on Papyrus: Form and Function, Content and Context*. Berlin: de Gruyter.
- Kaplony, A. (2008). What are those few dots for? Thoughts on the orthography of the Qurra Papyri (709-710), the Khurasan Parchments (755-777) and the inscription of the Jerusalem Dome of the Rock (692). *Arabica* 55(1): 91-112.
- Sijpesteijn, P.M. (2005). Checklist of Arabic Papyri, *Bulletin of the American Society of Papyrologists*, 42: 127-166. Updated version (accessed 1 October 2014): [http://www.naher-osten.uni-muenchen.de/isap/isap\\_checklist/index.html](http://www.naher-osten.uni-muenchen.de/isap/isap_checklist/index.html).
- Sijpesteijn, P.M. (2009). Arabic Papyri and Islamic Egypt. In Bagnall, R.S. (ed.), *The Oxford Handbook of Papyrology*. New York: Oxford University Press, pp. 452-72.