URL: http://nbn-resolving.de/urn:nbn:de:bsz:15-qucosa-201726

# The Corpus of Greek Medical Papyri and Digital Papyrology: new perspectives from an ongoing project

Nicola Reggiani

**[slide 2]** Three years ago, in Oslo, during one of the first official presentations of the project Corpus of the Greek Medical Papyri Online held at the University of Parma, now flown into the DIGMEDTEXT project, funded by the European Research Council (http://www.papirologia.unipr.it/ERC)[1], I advanced a range of trends, problematic in some ways, arisen by the digitisation of the Greek medical papyri (cf. N. Reggiani, *A Corpus of Literary Papyri Online: the Pilot Project of the Medical Texts via SoSOL*, "Antike Lebenswelten Althistorische und papyrologische Studien", hrsgg. R. Lafer und K. Strobel, Berlin-New York, De Grutyer, 2015, 341-52). It is indeed a corpus of very peculiar texts, which beside the well-known features of papyrus documents, implies issues connected to their special technical nature – which explains, if not justifies, their quasi-systematic exclusion from the usual documentary databanks and arises interesting questions on the digital edition of papyrus texts. Since then much has been endeavoured, both by the Parma team and on other frontlines, and it seems to me useful to present here a provisional appraisal, taking into account the current state of the art and stimulating further reflections on problems and difficulties not yet completely solved.

**[slide 3]** A first, more general issue was adapting the current platform for digitising papyri to the needs of a composite corpus containing also literary and "semi-literary" texts. The architecture of the *Papyrological Editor* was indeed designed to fulfil the needs of Greek and Latin documentary papyri: the XML skeleton based on a selection of the TEI standards already developed by Epidoc for the digital edition of ancient documents; the metadata taken from the HGV (later also from Trismegistos); the markup language Leiden+ designed to ease autonomous submissions by the community of the papyrologists. That issue has been absorbed by the very recent project *Digital Corpus of Literary Papyri*, started at the University of Heidelberg, of which the CPGM Online is a partner and a testing ground since the beginnings. The problems of how to encode textual or paratextual features almost absent from the documentary papyri (critical or diacritical marks, punctuation, marginal writings, graphical layout features, and so on), how to integrate the purely literary elements of a critical edition (textual variants, parallels, etc.), how to grasp the metadata from literary databases like the LDAB (**[slide 4]**), have been started up and are now in a process of proposal, discussion and solution by the team involved in building the new DCLP. Many encoding features already existing in Epidoc's XML schemes need to be recovered and possibly translated into Leiden+ language, others have to be created

[1] This paper was given within the framework of the said project (ERC-AdG-2013-DIGMEDTEXT [Prof. I. Andorlini], Grant Agreement No: 339828).

URL: http://nbn-resolving.de/urn:nbn:de:bsz:15-qucosa-201726

anew. Still open questions concern, for example, the markup of the different typologies of abbreviation, the possibility to search for such features, the possibility to link parallels and quotations to other online resources like the Perseus digital library or the TLG.

**[slide 5]** Another perspective I pinpointed to enhance the analytical power of a database of medical papyri is the application of a linguistic annotation on multiple layers. Annotating is a fundamental practice in the linguistic study of a corpus of texts: it allows to describe, record, interpret and analyse linguistic information at several levels, in which each layer corresponds to a particular category of relevant information. On the theoretical and practical correctness of treating Greek medical papyri as a true textual corpus I think there must be no doubt: a linguistic corpus is usually intended as a selection of sample texts representative enough of a language, and though the medical papyri at our disposal come from a random and incomplete selection, they can be considered as the entire reference population rather than as a sample of a larger group, so that linguistic annotation seems to me absolutely feasible. The basic annotation layer, related to the analysis of the parts of speech (the one also known as treebanking because it is usually represented with a tree graph) would allow to conduct an extensive lexical, phraseological-formulaic and syntactic analysis on the corpus, aimed also (but not only) at discovering styles and writing strategies specific of the medical texts, both literary and documentary: think only of the possibility to find out influences or interpolations between authors, or the presence of literary echoes in technical or documentary texts. To analyse in depth and comprehend the syntactic structure of texts would allow also to solve problems of interpretation, or even only to understand the exact meaning of a text (let us consider for instance the case of a schematic recipe as P.Oxy. 1088 (**[slide 6]**), where implicit verbs and asyndetic syntax would have to be made explicit). In the field of classical philology such linguistic analyses are now at a very advanced level, and in the meantime papyrology too has made important progress, visible now in the interesting project Sematia, conducted at the University of Helsinki by Marja Vierros and her colleagues (http://sematia.hum.helsinki.fi), aimed at facilitating the linguistic tagging of digitised documentary papyri (**[slide 7]**). There are still some open problems, like how to integrate annotation layers in the current SoSOL platform or how to treat fragmentary words or multiple alternatives of supplement, but we register undoubtful progress in this field too (**[slide 8]**).

**[slide 9]** There are also other annotation layers that might enhance our corpus furthermore. Layers of normalisation or regularisation (in morphology, syntax, vocabulary, phonology, orthography) are already provided by the SoSOL platform, in the form of "corrections" of various types. However, a more systematic analysis of such pieces of information might lead to design an annotation layer related to the different typologies of linguistic variations, to a systematic study of them, and to the definition of schemes and models which could bring, for example, interesting statistical

URL: http://nbn-resolving.de/urn:nbn:de:bsz:15-qucosa-201726

considerations on the doctors' cultural level. In this case too have been made some attempts to extract and analyse data concerning linguistic variants from what was already encoded in the existing databases (**[slide 10]** M. Depauw, J. Stolk, *Linguistic Variation in Greek Papyri: Towards a New Tool for Quantitative Study*, GRBS 55 (2015), 196-220; http://www.trismegistos.org/textirregularities): the advantages of an extensive integration of such an annotation layer in any database are apparent.

**[slide 11]** An annotation layer of "lemmatisation", that is the reduction of a declined or conjugated word to its original lemma, would prove essential in defining and analysing a specialised technical vocabulary like the one employed in the medical papyri, and it would represent an important bridge to connect the textual database to the related project Medicalia Online, consisting in an extensive lexical reference tool for ancient medical technical terms. From this viewpoint, a basic lemmatisation method is provided by the SoSOL platform through a glossary tool allowing to mark some keywords, of which brief definitions or explanations are inserted and displayed in pop-up boxes. A possible improvement of this type of annotation, with reference to the medical papyri, might lead to systematic connections to the lexical cards of Medicalia Online.

**[slide 12]**

**[slide 13]** A very interesting issue is the application to the medical papyri of the idea of transtextuality, defining the various possible relations among texts, as it has been recently analysed by Monica Berti: this does not relate only to the network of quotations and parallel passages, which we already considered earlier with reference to the possible improvement of the critical apparatus, but also to the aspect of fragment which very often the papyrus, be it literary or documentary, acquires. We can define the fragmentary character of the papyrus as a sort of "non-voluntary quotation", selected by the chance and by the material circumstances rather than by the will of an author. The transtextual link will be given in that case by the "virtual" existence of a hypertext (the original document, lost, more or less recoverable in a philological way) and the concrete one of a hypotext (our fragment). This perspective would bring innovative solutions to the current question of how to treat textual variants. A problem came up while defining the new encoding rules for the ongoing DCLP is indeed how to choose and represent the philological variants of a text: which critical edition to choose for reference, whether and how to represent the attested variants, how to consider a papyrus containing a textual variant unattested in the manuscript tradition, and so on. If we give up to consider a variant as a "wrong version" to be corrected and normalised, we can overcome the deadlock by looking at the set of variants as a network or a system, and by thinking the digital edition as a multitext, a place for a dynamic collation of several editions or versions, stratified in the time (**[slide 14]**). The Homer Multitext Project by the Center for the Hellenic Studies goes indeed along this path.

URL: http://nbn-resolving.de/urn:nbn:de:bsz:15-qucosa-201726

**[slide 15]** A last – but not least – aspect to be considered is the global one of the digital edition itself. The SoSOL platform, as is known, offers, beside a basic critical apparatus, the possibility to enhance the text with an introduction (called front matter), a line-by-line commentary, and a translation in one or more modern languages. Leaving aside the translation, however, a papyrus already published elsewhere is very seldom digitised in this complete way (an example of the most common appearance of a papyrus in the current databank: http://www.papyri.info/ddbdp/p.oxy;2;293). On the contrary, the medical corpus, though including published texts, aims at providing an overview as complete as possible for each text, where the reference to the paper edition is still unavoidable, but a first digital foretaste of it is given (http://www.papyri.info/editor/publications/45620/ddb_identifiers/103498/preview). This is, in my opinion, an important step towards those "complete" digital editions of which a recent seminar held at the University of Heidelberg is producing the first examples: unpublished or just described papyri published entirely and directly online, with the possibility of applying further improvements or corrections in the same way (http://www.papyri.info/ddbdp/p.got;;29dgtl). Despite some scholars still see in the paper edition an essential scientific goal, online publication is in fact one of the most stimulating perspectives of digital papyrology and, in general, of the entire field of digital humanities, because it combines the old problem of the "digital critical edition", object of long-standing philological debates, with the dynamic and innovative view of an open-access and collaborative platform like the Papyrologycal Editor. Likely we will never go as far as declaring obsolete glorious paper tools like the *Berichtigungsliste* or the *Sammelbuch*, but in projects just like the CPGM new spaces can open for entirely online, open-access, dynamic and interactive new textual editions.