Digital Papyrology II

Digital Papyrology II

—

Case Studies on the Digital Edition
of Ancient Greek Papyri

Edited by Nicola Reggiani

DE GRUYTER

Foreword

The present collective volume is conceived of as the ideal continuation of my monograph *Digital Papyrology*, which indeed appeared as Volume I with the same publisher. The two volumes are part of a project initially named *Beyond the Apparatus* intended to frame past and current issues surrounding the digital tools and methods that are being applied to papyrological research and scholarship. In the monograph, I tried to sketch the general outlines of electronic resources (bibliographies and bibliographical standards, metadata catalogues, virtual corpora, word lists and indexes, digital imaging processes, digital palaeography, information media, quantitative analyses, integrated workspaces, textual databases) in an attempt to define Digital Papyrology as a self-standing discipline that deals with meta-papyri, i.e. papyrus texts in the digital space. Accordingly, I argued that the ultimate purpose of Digital Papyrology is the digital critical edition of papyrus texts. The goal of the present volume is precisely to investigate this purpose, from the multifaceted viewpoints of the most advanced trends and projects in the field: namely, the deployment of platforms suitable for the encoding of proper digital critical editions of both documentary and literary Greek papyri and the development of quantitative analysis methods for the evaluation of the linguistic features of the texts.

In this challenge, I owe gratitude to my international colleagues and friends who have enthusiastically accepted to contribute with their invaluable experience in the field: in a rigorous alphabetical order, Rodney Ast (Heidelberg), one of the leaders of the *Digital Corpus of Literary Papyri* (whom I wish to thank for a linguistic revision of this Preface); Lajos Berkes (Berlin), member of the *Papyri.info* editorial board and author of several born-digital editions of documentary papyri; Isabella Bonati (North-West University, Pochetsfroom, South Africa), soul of the lexicographical project *Medicalia Online*; Giuseppe Celano (Leipzig), co-editor of *The Ancient Greek and Latin Dependency Treebank*, with his long-standing experience in treebanking and morphological annotation of classical texts; Holger Essler (Würzburg), DCLP partner and architect of digital projects about linguistic annotation (the *Annotated Philodemus*), image alignment and automated character recognition in the Herculaneum papyri (*Anagnosis*); Massimo Magnani (Parma), who kindly agreed to bring a brilliant classical philologist's viewpoint to the evaluation of the issue at stake; Joanne Stolk (Ghent), co-editor of the *Trismegistos* database of *Text Irregularities*, with her strong experience in linguistic variation in the papyri and its digital treatment; Marja Vierros (Helsinki), who launched (and manages) the path-breaking platform *Sematia* aimed at facilitating linguistic annotation of the papyri.

On my side, I wish to acknowledge the fact that the volume stems from the project "Online Humanities Scholarship – A Digital Medical Library of Ancient Texts" (DIGMEDTEXT: http://www.papirologia.unipr.it/ERC), funded by the European Research Council (Advanced Grant Agreement no. 339828) at the University of

Parma (2014–2016) and directed by Professor Isabella Andorlini, to the grateful memory of whom this volume is dedicated. This statement is not a matter of pure bureaucracy. The DIGMEDTEXT project, primarily aimed at creating a database of the Greek medical texts on papyrus, has been the breeding ground for more general – theoretical, methodological, and technical – reflections about linguistic papyrological phenomena and their electronic treatment, as well as about the digital critical edition of the papyri themselves. It is my hope that the entire papyrological community, and in general all scholarship interested in such topics, will enjoy the results reached in the past years, and that discussion and development may continue further in the future.

Parma, January 10, 2018   Nicola Reggiani

Contents

Part 1: **Platforms Between Theory and Practice**

Nicola Reggiani

The *Corpus of the Greek Medical Papyri* and a New Concept of Digital Critical Edition

1 Defining and shaping a digital critical edition

Traditionally and basically, a critical edition of a text is the printed output of a philo-logical work, i.e. the process of reconstruction of a textual archetype (the 'source') among different variants, aimed at reproducing the original text as most exactly as possible, or, in other terms, as the fixed representation of a scholar's more or less trustable opinion on that text. Accordingly, and rather intuitively, a digital critical edi-tion should be defined as the digital output of a philological work. We will see what a "digital output" involves in methodological and epistemological terms but, to start, it must be noted that traditionally a digital critical edition is regarded as the digital transfer of a printed critical edition. Sometimes, this process regretfully gets rid of the attribute 'critical', so that we have digital editions or textual corpora deprived of ap-paratus criticus and therefore 'uncritical', as in the well-known cases of the *Thesaurus Linguae Graecae* or of the *Perseus Digital Library*. This treatment presents encoding advantages, since one reference edition is chosen and digitized, but also huge disad-vantages in terms of usability, because search and analysis functions are limited to the chosen text, without consideration, e.g., for textual variants, alternatives or dif-ferent editorial solutions.[1] Somewhat 'hybrid' editions try to save the *constitutio textus* (the restitution of a text as close as possible to the supposed original) alongside the recording of variant readings: for example, the former *Duke Databank of Documentary Papyri* with the spelling variants (as written on the original papyrus) embedded within the 'normalized' text with special markup.[2] A fairer transfer process preserves the apparatus criticus, which is usually displayed in a way that resembles the printed edition. The simplest examples are PDF editions (either scans of paper samples or born-digital files like the publications of the *PHerc* project),[3] the most articulated ones are the digital editions available at the *Papyri.info* platform, where critical annota-tions, encoded as inline XML markup elements, are processed and displayed in an

---

1 See already DEGANI 1992, and more recently MAGNANI 2008, 135–7; also M. Magnani in this volume.
2 Cf. REGGIANI 2017, 215–7.
3 Cf. REGGIANI 2017, 176.

apparatus of print-like format, stressing the distance between the 'correct' text and alternatives, variants, actual textual features.

Fig. 1: PSI XV 1510, medical catechism on anatomy, III cent. AD: printed edition and digital edition at http://litpap.info/dclp/64024.

This traditional view is being challenged as rather uncomfortable by the development of digital technologies in the ancient studies, as well as by an increasing concern for the actual testimonies and the process of textual tradition: we may define it as a sort of 'phenomenological' approach. Digital projects like the *Homer Multitext Project* (HMT) or the *Leipzig Open Fragmentary Texts Series* (LOFTS) started envisaging a different approach to textual criticism, in deploying a text that is in fact a multitext, a fluid and dynamic network of multiple editions aligned to each other (by means of a URN architecture) rather than a traditional fixed structure of text and apparatus criticus.[4] In this framework, the uneasiness of texts that are felt not being completely suitable for a 'traditional' critical edition (e.g. oral Homeric poetry,[5] fragmentary

---

4 A multitext is basically a dynamic collection of multiple critical editions, a network of versions with a single root. As Monica Berti described it, "[i]t produces a representation and visualization of textual transmission completely different from print conventions, where the text that is reconstructed by the editor is separated from the critical apparatus that is printed at the bottom of the page. [...] It] allows the reader to have a dynamic visualization of the textual tradition and to perceive the different channels of both the transmission and philological production of the text that is usually hidden in the static, concise, and necessarily selective critical apparatuses of standard printed editions. Producing a multitext, therefore, means producing multiple versions of the same text, which are the representation of the different steps of its transmission and reconstruction, from manuscript variants to philological conjectures" (BERTI forthcoming, 4). Cf. REGGIANI 2017, 266 ff.

5 The HMT project concept results from the statement that the Homeric textual evidence does not comply with the traditional philological view of textual variants stemming from one archetype, since

sources[6]) merges with the new capabilities of digital infrastructures, which offer much more dimensions than printed paper. Hypertext is a new writing space, to which editors have to adapt the texts:[7]

> [o]nce we are able to overcome the physical limits of printed editions by joining together variants and conjectures referring to the same texts, it also becomes possible to look at the texts from a new and broader perspective, with possible consequences for our knowledge and comprehension of them.[8]

Thence, an unavoidable fact:

> [w]e need to move in the direction of digitally conceived and initiated types of information and away from mopping up information from print sources.[9]

As it has been put very effectively, the hypertext architecture is challenging the *Urtext* model,[10] and it paves the way for exploring the possibilities of "holistic" models where editorial choices are superseded by an interactive network of all extant data, with potentially infinite information layers.[11] Perhaps, the model that better describes this ideal condition is an ontology design:

> an ontology is the most suitable solution to represent critical editions of ancient texts for two main reasons: first, we want to be able to link different kinds of resources […] that have in common the possibility of being referred to via URIs, which is one of the principles of the Semantic Web; second, information contained in critical editions constitutes a layer of interpretation and a description of relations about texts that is important to keep clearly distinct from the texts themselves. Indeed, the use of stand-off metadata encoded within ontology allows us to express an open-ended number of interpretations, whereas a markup-based solution would not make this possible due to obvious reasons of overlapping hierarchies.[12]

---

a true original Homeric text never existed (cf. BIRD 2010): a somehow "agnostic" (BODARD – GARCÉS 2009, 96 n. 31) environment where all witnesses are transcribed and juxtaposed, without preference for any of them. See M. Magnani's chapter in the present volume for a critical view of this idea.

6 Ancient fragments are characterized by a high level of textual complexity, in the relationships among the actual text in which they are embedded, its critical edition (interpretation), the original source (attribution), the quoting source (witness), etc.: cf. BERTI forthcoming.

7 Cf. BOLTER 1991; REGGIANI 2017, 263 ff.

8 ROMANELLO – BERTI – BOSCHETTI – BABEU – CRANE 2009, 165

9 BAGNALL – GAGOS 2007, 74.

10 BOLTER 1991.

11 Cf. BODARD – GARCÉS 2009

12 ROMANELLO – BERTI – BOSCHETTI – BABEU – CRANE 2009, 158. An ontology is a formal definition of types, properties, and interrelationships of the entities belonging to a certain domain of knowledge. In other words, it compartmentalizes the variables needed for some set of computations and establishes the relationships between them.
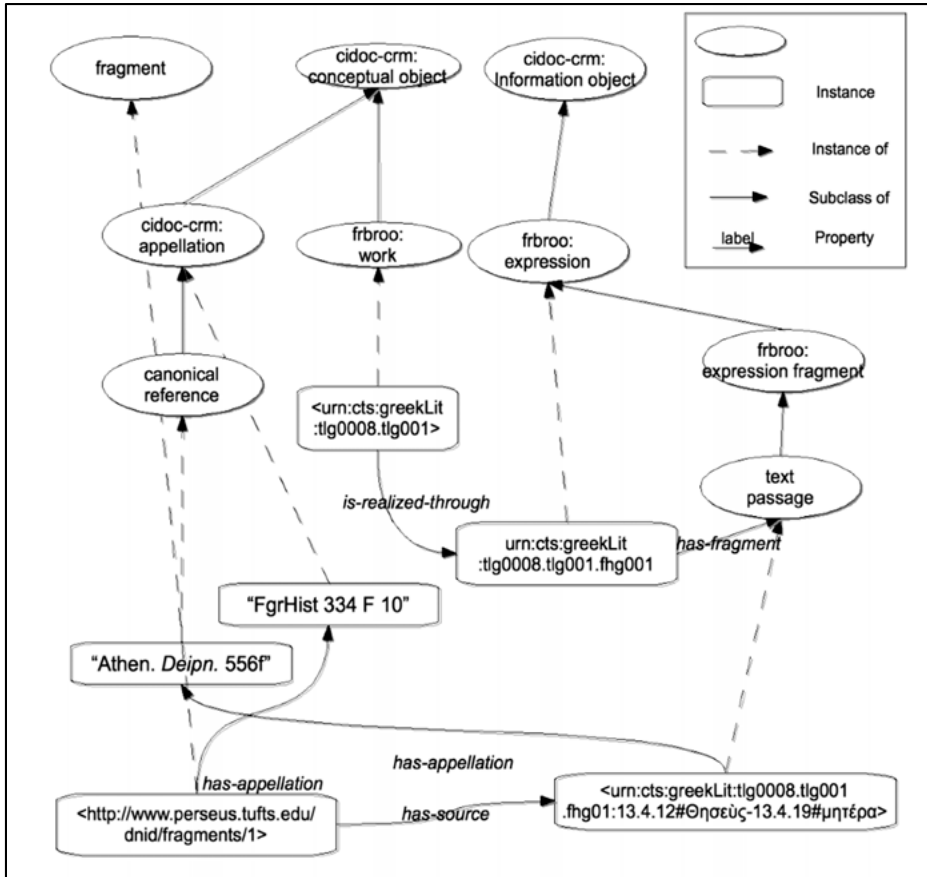
Fig. 2: A sample ontology model (from Romanello – Berti – Boschetti – Babeu – Crane 2009, 167).

2 Papyrology: philology in flux

Papyrology is, in its more essential core, all about providing trustable critical editions (and commentaries) of papyrus texts.[13] Though projected towards a broad historical and cultural evaluation of the textual data,[14] it is intimately a philological discipline:[15] no one can deny that without texts there would exist no Papyrology. Yet it is a very peculiar philological discipline, since it is well aware of the fluidity of its objects of

---

13 Cf. Youtie 1963, 22–3.
14 Cf. e.g. Bagnall 1995.
15 Cf. Hanson 2002, 196; Schubert 2009, 197.

study:[16] texts are continuously published, updated, collected, revised, corrected, emended, republished, and there is hunger for resources that can help handling an overwhelming amount of primary data.[17] It is – to borrow the successful concept that Zygmunt Bauman launched to emphasize the fact of change in the modern times[18] – a 'liquid' philology, for which digital environments seem extremely fitting; in particular, collaborative platforms like SoSOL seem the most suitable incarnation of this complex and fluid editorial workflow.[19]

Moreover, Papyrology has always been facing an adventurous textual situation, having to cope with fragmentary and unique texts and idiosyncratic utterances, and has developed a remarkable interest in the scribal and material phenomenology of textual features and transmission, which affects consistency in treating the wide series of textual fluctuations occurring in the papyri. Indeed, while philological analysis would gladly treat fluctuations as deviations from a standard archetype (i.e. mistakes or, more gently, variants) and normalize them in a reconstructed critical edition, they actually bear significant socio-cultural relevance and are of fundamental importance from the viewpoint of the phenomenology of the papyrus texts, its interpretation, and ancient writing culture in general. In other words, very often fluctuations are not used to reconstruct a text but to investigate relevant socio-cultural phenomena. Accordingly, the papyrologists' behaviour towards such textual flavours is twofold, and generates a wide variety of editorial inconsistencies that affect printed editions as well as digital databanks.

As to the latter, the issue at stake is not only critical agreement or scholarly standards, but also (as hinted above) the usability of the tools themselves, in terms of searching and encoding. The best example, from my own experience, is the case of the word ἑρμηνεία, which often occurs in the papyri in the iotacistic form ἑρμηνία. The spelling 'variant' is treated differently in the printed editions, being sometimes 'regularized' in the apparatus, sometimes not, generating textual inconsistencies even within the very same text.[20] In BGU I 326, ii 15 ἑρμηνία is printed without apparatus notes, and it is reproduced in the databank as such; in the same text, at l. i 1 the same word is supplied as ἑρμηνεί]α (following the 'standard' form) in the database, while all the printed editions (after the *ed.pr.*: Chr.M. 316; Sel.Pap. I 85; FIRA[2] III 50; Jur.Pap. 25) keep the 'variant' in the lacuna too. Another 'classical' case is that of the

16 Cf. Youtie 1963, 27–32; Hanson 2002, passim; Schubert 2009, 212–3.

17 As I pinpointed in Reggiani 2017, 2–6, this is the basic *raison d'être* of Digital Papyrology.

18 Cf. e.g. Bauman 2000; 2007; 2011.

19 On the collaborative structure of the database cf. Reggiani 2017, 232 ff. All editorial interventions are kept recorded in a "History" log, which is available to every user: see L. Berkes' article in this same volume for a screenshot of a sample editorial history on *Papyri.info*.

20 Cf. Reggiani 2018a. Some remarks on the inconsistent treatment of iotacism can be found also in J. Stolk's and M. Vierros' contribution to this volume.

verbal forms of γίγνομαι, which becomes γ(ε)ίνομαι in the *Koine* Greek.[21] The latter forms are indeed treated as the standard in most of the papyrus editions, and therefore are not 'regularized' as variants,[22] but this is not always consistent: editorial regularizations do occur, seemingly only when the verb is affected also by iotacism, often in compounds.[23] On the other hand, we do find the classical Greek forms not being regularized as well,[24] which increases the uneasiness of anyone who would like to perform effective searches in the digital textual corpora. With the further developments of the Greek language, the situation is even more complex: for example, the general shift from dative to genitive in the later (Byzantine) instances of the language of the papyri[25] leads to further editorial inconsistencies. In BGU XIII 2332,20 (AD 375), for instance, ὑπάρχω + genitive (μου) is regularized in dative (μοι) according to the classical use,[26] whereas in SB XVIII 13947,15 (AD 507) ὑπάρχω + dative (μοι) is regularized in genitive (μου) as if the latter was then the correct form.[27] One must be aware of any possible spelling or syntactic combination to perform trustable textual searches.[28]

As is apparent, papyrus texts carry a cognitive complex that is often hard to fit into printed editions and may find its better representation in the digital space, where the objects of study undergo a process of dematerialization. I have already argued that the development of Digital Papyrology, in its treatment of computerized information about papyri, produced the effect of working on the virtual representation (*avatar*) of the papyri themselves, which turn to be meta-texts,[29] in the terms already envisaged by Traianos Gagos as early as 1998:

> In this new era of papyrological research, we cannot speak of a collection of papyri alone, but also of a collection of electronic files, data, metadata and digital images:[30]

---

21 Cf. Depauw – Stolk 2015 and J. Stolk's chapter in this volume.

22 A quick survey of a sample search in *Papyri.info* can give a global idea of this trend: http://papyri.info/search?STRING1=γεινομ&target1=TEXT&no_caps1=on&no_marks1=on&STRING2=NOT+γιγνο&target2=TEXT&no_caps2=on&no_marks2=on.

23 παραγ{ε}ινεται *l.* παραγίγνεται in BGU XVI 2651,6; γείνεσθαι *l.* γίγνεσθαι in Chr.M. 172,i,15; καταγειν[ο]μ[αι] *l.* καταγίγνομαι in P.Bodl. I 17,i,9; παραγεινομαι *l.* παραγίγνομαι in P.Haun. II 22,5; περιγεινομένων *l.* περιγιγνομένων in P.Stras. VIII 772 passim. Note the double possible regularization γίγνεσθαι or γενέσθαι advanced for γείνεσθαι in P.Col. X 280,13.

24 Another sample search: http://papyri.info/search?STRING1=γιγνομ&target1=TEXT&no_caps1=on&no_marks1=on.

25 Cf. Stolk 2015b.

26 For more similar cases cf. Stolk 2015a, 85 ff., and 2015b.

27 Cf. Depauw – Stolk 2015, 213. See also Stolk 2015a, 93.

28 On these topics cf. Reggiani 2018a, 2018b, 2018c, and J. Stolk in this volume.

29 Cf. Reggiani 2017, 260 ff.

30 Gagos 2001, 516.

The availability of huge amounts of information in fully searchable textual form with accompanying images through these new media is altering drastically the definition of what constitutes a 'text', the way we experience reading it and, ultimately, the plurality of messages a text can offer to one or more readers. The new methods of presenting text with marked up images and the simultaneous availability of a variety of other research tools within the same electronic environment give us new ways of visualizing and approaching a given text. An edited text is no more a static, isolated object, but a growing and changeable amalgam: the image allows the user to look critically at the 'established' text and to challenge continuously the authoritative readings and interpretation of its first or subsequent editors.

Furthermore, the simultaneous access to and study of thousands of texts and their images that could be as far apart as a millennium, in a single search and through the same medium, has the potential to challenge our established notions of the 'messages' a text carries within itself, its textuality and intertextuality [...]. As Roland Barth [*sic*] explains: 'Any text is an intertext; other texts are present in it, at varying levels, in more or less recognizable forms: the texts of the previous and surrounding cultures. Any text is a new tissue of past citations. Bits of codes, formulae, rhythmic models, fragments of social languages, etc. pass into the text and are redistributed within it, for there is always language before and around the text'. In one or another way, papyrologists have always recognized the "intertextuality" of the Greek papyri from Egypt, because of the multicultural and multi-ethnic environment in which these texts were born. The development of the new electronic media in our field and the capability to establish these cross-links – or these intertextual signifiers, so to speak – on the linguistic, cultural and historical level through the interaction of multiple texts, images and a variety of related tools places the notions of textuality, intertextuality and metatextuality on a new (electronic) platform which, in turn, becomes part of these notions as the 'carrier', 'interpreter' and 'distributor' of these texts.[31]

The concept that Digital Papyrology redefines the notion of 'papyrus' is embedded in the consideration that

these media, when used within a wider intellectual perspective as a cognitive tool for research and instruction and not only as a pragmatic medium that can 'do certain things for us', can challenge and redefine notions of 'text' and textuality.[32]

After realizing that we are coping with enhanced papyri that are in fact 'meta-papyri', we need to reshape the digital edition in accordance with the nature of the papyrological digital data as autonomous intellectual objects (following the definition of what is 'data' for the humanists according to Owens 2011),[33] and the possibilities offered by the electronic meta-space.[34] There is a momentous chance to see the digital document not as the mere, more or less complete reproduction of a printed critical

---

31 Gagos 2001, 514–6.

32 Gagos 2001, 515 n. 8.

33 At the same time constructed artefacts, being created by people, and interpretable texts, they "can hold the same potential evidentiary value as any other kind of artifacts".

34 See also the observations by M. Magnani in this volume.

edition, but as a quantum particle of a fluid universe of text transmission. This 'dispositive' – in foucaldian terms[35] – may find a suitable representation through the abovementioned ontology design, where we do not have to decide what is 'regular' or 'normal' and what is a 'secondary' reading, but can create an interconnected network of aligned versions, which represent different possible layers of textuality:[36]

> [o]nly with a comprehensive understanding of the content and assumptions of the traditional hughly-evolved critical apparatuses will we make the right strategic decisions for the future of textual scholarship.[37]

Philology tends to overcome any textual fluctuation in favour of a reconstructed text that be as closest as possible to the 'original' source, but documentary papyri are actually the original source of themselves (any critical interventions being configured as the reconstruction of an imaginary archetype), while literary and paraliterary papyri present more complex issues, as introduced below. They are therefore among the best text typologies suitable for exploring new ways of conceiving digital critical editions.

3 The medical papyri: special technical needs of a special technical corpus

Within the framework sketched above, the *Digital Corpus of the Greek Medical Papyri* project[38] proved pathbreaking in applying the notion of digital edition to literary and paraliterary papyri, previously excluded from *Papyri.info* and object of very specific and isolated projects (CPP, THV etc.). The project stemmed from Isabella Andorlini's lifelong interest in the medical papyri and from her own challenge to collect them in

---

35 Lamé 2014 describes this idea (with reference to ancient epigraphs) through Foucault's philosophical concept of *dispositive*: the message of the text-bearing object can be completely understood in relation with a complex network of many other heterogeneous pieces of information. The ultimate purpose is "to digitize also the network that connected those information systems, instead of digitizing each individually".

36 The platform *Sematia*, discussed by M. Vierros in this volume, is a nice example of how the transcription of the actual papyrus text can be aligned to a 'regularized' layer of the same text, so that any possible information is kept in an interactive way.

37 Damon 2016, 218.

38 With DCGMP I refer to the whole digital corpus of the Greek medical papyri as resulted from the work of the DIGMEDTEXT project mentioned in the Introduction to this volume. The title *Corpus dei Papiri Greci di Medicina Online* ("Online Corpus of the Greek Medical Papyri") refers to the first stages of the project. Bibliography: Andorlini – Reggiani 2012; Reggiani 2015; 2016a; 2017, 273–5; Andorlini 2017; Bertonazzi 2018a, 24–9; Reggiani 2018b; http://www.papirologia.unipr.it/CPGM.

a uniform and homogeneous corpus.[39] Her first steps went towards the printed medium,[40] but she soon realized the strong potentials of *Papyri.info* to host dynamic papyrological editions,[41] and her project later became one of the leading pilot test cases of the rising *Digital Corpus of Literary Papyrology*[42] in envisaging new technical and theoretical strategies for the encoding of literary and paraliterary texts, eventually awarded with an ERC advanced grant (http://www.papirologia.unipr.it/ERC).

Medical papyri are technical texts: they have been conceived to convey a technical knowledge, i.e. theoretical and practical specialized information at the same time – a knowledge that is, in turn, mirrored and refracted in the different written genres encompassed by the corpus.[43] The importance of medical technical skills is apparent, and not only for health reasons (think of Galen's instructions to the patients so that they can choose the best doctor after an enquiry on his skills):[44] one might recall P.Oxy. I 40 (+ BL I 312, V 74, VI 95; Oxyrhynchus, II cent. AD), a copy of the report of a court judgement where a public doctor claims for immunity from some liturgies, and the judge, after a rather witty remark, requests a scientific proof of his assertion.[45] The importance of written text for this education is stressed as earlier as in the Hippocratic corpus: "I consider the ability to evaluate correctly what has been written as an important part of the art" – says the author of the *Epidemics* – "He who has knowledge of it and knows how to use it will not commit, in my opinion, serious errors in the professional practice" (*Epid*. III 16 = III 10,7 ff. L.). In fact, the transmission of this knowledge was carefully carried out through a specialized education, which was based on oral teachings later entrusted to written supports. In the introduction to the treatise *On his own books*, Galen himself explains how in the context of the oral lesson one used to take written notes, thence moving to the publication of *memoranda*, the *hypomnemata* of the lessons heard.[46]

Stemming from both the knowledge of oral teaching and the know-how of practical records and individual experience, every medical writing is not a fixed book but a tool in flux: the older treatises are annotated, commented, collated often against annotated and commented copies,[47] transcribed with additions, corrections, and updates; the collections of personal notes on clinical cases, therapies or remedies are

---

39 Cf. Reggiani 2018d.

40 Cf. Andorlini 1997a.

41 Cf. Andorlini – Reggiani 2012, 138–9; Bagnall 2012, 4.

42 See the chapter by R. Ast and H. Essler in this volume.

43 Cf. Andorlini 1993; Reggiani 2018e.

44 Cf. Nutton 1990.

45 On official examinations of physicians see Reggiani 2018f.

46 Cf. Nutton 1972; Nieddu 1992, 555–7; Andorlini 2003, 14.

47 On the collation of annotated copies, always according to Galen's words (*In Hp. Off*. III 22 = XVIIIb 863,14–865,5 K.; *In Hp. Epid*. II 8 = XVIIa 634,3–7 K.), cf. Andorlini 2003, 15, who recalls (note 15) the story of Mnemon, who took the third book of Hippocrates' *Epidemics* from the library of Alexandria

constantly revised on the ground of practice; prescriptions are transcribed, exchanged, collected, gathered in the *receptaria* and passed down; handbooks of different typologies are used to teach again, and so on, keeping on the written support traces of every stage of transmission and use.[48]

Such texts are further characterized by intertextual and transtextual connections: references, quotations, more or less literal parallels, are another key to understand and contextualize the matter at the best. The recipes, and their collections known as *receptaria*, find inspiration in the pharmacological treatises and are further enriched by the doctors' personal practice and by references and quotations from different medical sources; the questionnaires are connected to the literary tradition of the *Definitiones medicae* (see below). These are by no means stemmatological relationships between ascendants and descendants: it is a fluid knowledge undergoing continuous changes, updates, adaptations, much influenced by oral teaching and actual practice. Accordingly, the very textual data interweave with a huge panel of textual devices, which contribute to articulate an expressive network that is essential to the medical writing itself, to its transmission, to its learning, and to its practical use: therefore, they deserve a particularly careful consideration. Critical and diacritical marks, punctuation, graphical and layout features, technical terms and formulae, literary or sub-literary references or echoes, marginal annotations – to cite the most outstanding devices – form a complex interplay that cannot be separated from the text itself, nor – even more – ignored, without compromising the correct interpretation of the evidence. Rigid definitions of philological variants do not really apply, as well as the treatment of linguistic variants can be more complex than the simple application of regularization markup tags, which categorize a 'standard' (not to say 'correct') and a 'deviant' version of a word.[49]

The inadequacy of the traditional philological/stemmatological model to represent in full the textual features of these complex and fluid technical writings has already been pointed out by Ann Hanson,[50] who advanced an "accretive model of composition" to provide a suitable description of the phenomenon. In David Leith's words,

> [t]he textual tradition of compilations of this sort was highly fluid, and we should not conclude that they represent exactly the same text.[51]

---

and brought it back with the marginal addition of marks indicating clinical histories, traced with dark ink and big letters, in imitation of the original handwriting. Cf. also BONATI 2016b, 63–4, and see below.

48 Cf. ANDORLINI 2003; REGGIANI 2018e and 2018g.

49 Cf. REGGIANI 2018a for further details, and see below.

50 HANSON 1997.

51 D. LEITH, P.Oxy. LXXX 5239.

Digital tools offer now the best solutions to face this challenge, which Isabella Andorlini herself envisioned at the very beginning of the *Corpus dei Papiri Greci di Medicina* project, a primary focus of which was

> la nuova attenzione rivolta alle problematiche editoriali dei testi studiati nella complessità dei rapporti con le fonti rispetto alla tradizione conosciuta.[52]

4 Envisioning digital critical editions of (medical) papyri

As I anticipated above, a digital critical edition can be defined as the digital output of a digital philological work. This has a vital outcome in terms of data encoding. Indeed, encoding data involves a digital critical workflow that takes the features of the original texts (and of the printed editions) to adapt them to the digital medium. It requires a thorough philological work, namely a digital philological one, where the digital papyrologist is – to paraphrase Youtie's well-known definition – an artificer of data (in the abovementioned, intellectual meaning of 'data'). Any information taken from the text or from previous editions becomes data (or metadata, i.e. 'data about data'); and even when encoding a print-published edition, one should check carefully the original text to avoid possible inconsistencies and ambiguities inherited by the previous editors, so that the 'liquid' editorial flux goes on.

Moreover,

> [e]ncoding fragments is first of all the result of interpreting them, developing a language appropriate for representing every element of their textual features, thus creating meta—information through an accurate and elaborate semantic markup. Editing fragments, therefore, signifies producing meta—editions that are different from printed ones because they consist not only of isolated quotations but also of pointers to the original contexts from which the fragments have been extracted. On a broader level, the goal of a digital edition of fragments is to represent multiple transtextual relationships as they are defined in literary criticism [...]. Designing a digital edition of fragments also means finding digital paradigms and solutions to express information about printed critical editions and their editorial and conventional features. Working on a digital edition means converting traditional tools and resources used by scholars such as canonical references, tables of concordances, and indexes into machine actionable contents.[53]

Therefore, "encoding a text is an interpretive act"[54] by itself: on the one hand, the encoder (the digital papyrologist) must employ as much criticism and careful discern-

---

52 ANDORLINI 1997a, 19 (cf. *ibid.*, 21–2)
53 BERTI forthcoming, 2.
54 OWENS 2011.

ment as possible in order to give the papyrological object its correct digital representation. On the other hand, one must be aware of the fact that the digital medium has different requirements than the printed one. While philology is a way of *describing* a text to interpret it as a stable source, a phenomenological approach is a way of *representing* a text in all its components, to describe and understand the underlying semantics. Therefore, when we choose to overcome the inadequacies of a traditional critical edition in favour of the digital multi-space, we must keep in mind the following three fundamental requirements:

- standardization (adapting to the digital medium means to follow its strict rules);[55]
- semantic representation (which may differ from the traditional philological representation, as we will be noticing below);
- usability (in terms of data access, searching and developing options).

Data and metadata can be encoded and used as different, yet interconnected (aligned) information layers.[56] An XML annotation markup seems to be the best encoding strategy, since it has a consolidated background in the TEI/EpiDoc system that has already been adapted to the papyrological requirements,[57] providing a standardized and standardizing framework, a semantic annotation, and powerful search options through XPath and XQuery querying languages.[58] It also allows for any kind of final rendering by means of customizable transformation languages (XSLT). Alignment among layers can be achieved by deploying a CTS URN architecture, which is useful to give unique identifiers to each element and to avoid overlapping hierarchies, especially in linguistic annotation. Annotated layers can be stored in a GIT repository so that open access and collaboration are granted. Some layers already exist in the SoSOL infrastructure (metadata, introduction and commentary, translation, annotated text); more can be envisioned, for example, on the ground of Gérard Genette's textual theory, which describes all possible relations among texts and which has already been claimed as the privileged interlocutor of the complex textual 'dispositive' of papyrus texts.[59]

---

55 This is indeed a key issue in Digital Papyrology (cf. REGGIANI 2017, passim) as felt by the very first 'fathers' of the papyrological databases (cf. TOMSIN 1970, 476).

56 I started envisaging this strategy for the medical corpus in REGGIANI 2015, where I sketched some possible annotation layers (the article stems from a conference paper delivered in 2012, at the very beginnings of the DCGMP project). I revised my argument in REGGIANI 2016a.

57 See the overview discussed by J. Stolk in this volume. In the following pages, I will be referring to the online Leiden+ guidelines at http://papyri.info/docs/leiden_plus.

58 See the query cases mentioned by M. Vierros and G. Celano in this volume.

59 "This 'intertextuality' of the text is what G. Genette would call 'transtextuality'. It is not, perhaps, accidental that postmodern theories on language and 'text' developed more or less at the same time with the spread of the electronic media" (GAGOS 2001, 515 n. 8). Cf. GENETTE 1992. I outline the possible exploitation of Genette's textual theory in relation with the complex textuality of Greek medical papyri in REGGIANI 2018c and 2018e.

Without any presumption of emulating Herbert C. Youtie, who provided the standard outline of a 'canonical' papyrus edition,[60] what follows is an attempt of systematizing the extant strategies for encoding a digital papyrus edition, with some suggestions for possible further improvements. The past work on the medical papyri provided the most complex and intriguing cases, but the same recommendations can apply to simpler cases too, as well as to documentary papyri of any sort.

4.1 Metadata and bibliography

Papyrus metadata (i.e. contextual information about texts: chronology, provenance, etc.) are currently stored in digital catalogues like the *Heidelberger Gesamtverzeichnis* (HGV) for the 'documentary' texts, the *Leuven Database of Ancient Books* (LDAB) and the *Mertens-Pack³* (M-P³) for the 'literary' ones, *Trismegistos* (TM) for both, and some more specialized ones like the *Corpus of Paraliterary Papyri* (CPP) or *Synallagma*. Similarly, digital bibliographical repositories exist, namely the *Bibliographie Papyrologique* (BP) and *Trismegistos* bibliographies.[61] *Papyri.info* and the DCLP currently include metadata from (respectively) HGV + TM and from LDAB,[62] and point to BP records as well as to some more resources (e.g. *Synallagma*). A digital critical edition of medical papyri should of course extend this feature to include the *Mertens-Pack³* (in its specific *Medici et medica* section)[63] and possibly to envision some digital version of Marganne's and Andorlini's printed catalogues of medical papyri.[64] Contextualization is indeed fundamental:[65]

> [l]o studio del manufatto e una sua corretta collocazione cronologica sono informazioni essenziali, che possono interferire con le ipotesi di attribuzione dei contenuti, sia per il rapporto con gli autori noti, sia per un'adeguata impostazione dell'indagine sulle fonti e sugli anelli della tradizione indiretta. La provenienza del reperto papiraceo può, nei casi in cui gli elementi archeologici siano conosciuti, conservare dati preziosi sul contesto in cui inserire le farine di produzione libraria antica, e sui livelli della sua divulgazione in Egitto (centri di diffusione legati alle vie dell'insegnamento e della pratica della disciplina; biblioteche templari, scuole mediche specializzate): l'attenzione ai luoghi accertabili di ritrovamento dei reperti ci permette di delineare il *milieu* culturale in cui libri di questo genere furono prodotti, o semplicemente letti, da fruitori professionisti e da gente colta con qualche interesse per i temi della salute.[66]

---

60 Youtie 1963, 22–3.

61 On these resources cf. Reggiani 2017, 39 ff. (catalogues) and 14 ff. (bibliographies) respectively.

62 See R. Ast and H. Essler in this volume.

63 Cf. Marganne – Mertens 1997 and the online resources cited in Reggiani 2017, 34.

64 Marganne 1981a; Andorlini 1993.

65 See also M. Vierros' remarks about metadata of documentary papyri in her article for this volume.

66 Andorlini 1997a, 21.

Moreover, the standardizing potential of digital metadata[67] would be a nice ground to deal with the problem of the definition of textual genres or typologies, and to face the challenge of a categorization, an issue that is well framed – from the medical viewpoint, in the wake of Isabella Andorlini – by Francesca Bertonazzi in the following words.

Classificare i papiri per tipologia non è solo un mero esercizio erudito o, peggio, sterilmente matematico nel senso deteriore del termine: al contrario si configura come un'indagine che può gettare luce sul contesto di composizione e d'uso del testo, e non di rado può agevolare la sua ricostruzione filologica e l'interpretazione esegetica. L'attività non è priva di rischi: un primo problema è sottolineato da quanti mettono in guardia dalla rilevanza statistica dei dati che possono essere desunti dai papiri, che sono inevitabilmente vincolati ai ritrovamenti, al tipo di descrizione fornita dal primo editore, dal tipo di classificazione operata nei primi studi sul testo. Il primo ostacolo, per così dire, è dunque di natura extratestuale, ovvero risiede nella mera quantità di papiri appartenenti a una data tipologia: anche se la maggior parte dei papiri medici afferissero al genere, e.g., del trattato, non per questo si dovrebbe concludere che il trattato fosse il genere più praticato in ambito medico nell'Egitto greco-romano. Un secondo problema, di tipo intratestuale, risiede nella tipologia stessa del documento, che spesso non appartiene in modo netto all'uno o all'altro tipo di testo: "Chi si è occupato anche solo marginalmente della interpretazione di frammenti di papiro a contenuto 'medico', avrà constatato come una delle difficoltà più evidenti è quella del riconoscimento e della definizione del genere testuale, del tipo di opera cui appartennero brani parziali di scritti oggi in larga parte perduti. Una difficoltà dovuta, oltre che alla casualità e alla precarietà del reperto papiraceo, anche alla organizzazione stessa delle opere a contenuto medico, teorico o specialistico che fosse: il riconoscimento di soggetti e termini medici è da solo insufficiente per dirci qualcosa di più preciso sull'impostazione dell'opera originaria, in quanto le singole nozioni tecniche ricorrevano in settori diversi della disciplina, e potevano essere esposte o discusse a livelli di approfondimento e di concettualizzazione anche molto distanti tra loro" [ANDORLINI 1997b, 159].

In quest'ottica, lo studio del *corpus* offre alcuni casi interessanti di testi a mezzo tra l'una o l'altra tipologia (come P.Oxy. 2.234 + 52.3654,92 tra il catechismo e la raccolta di prescrizioni), oppure di informazioni testuali insufficienti a distinguere con precisione l'appartenenza tipologica (come in P.Oxy. 74.4973: il testo potrebbe riguardare la veterinaria come la fisiognomica), o ancora di testi che pur rientrando nella categoria 'lettera', possono avere natura documentaria (come MPER 13.6 e GMP 2.10, lettere redatte da medici, e P.Mert. 1.12, lettera a un medico) oppure letteraria (P.Oxy. 9.1184 raccoglie varie lettere di Ippocrate).

Un terzo problema, di ordine linguistico, risiede nella terminologia moderna utilizzata per classificare i testi: non di rado si è avvertita la necessità di puntualizzare le varie accezioni di 'etichette linguistiche' attribuite a generi antichi: "[n]el classificare la ricettazione nei papiri ho volutamente differenziato l'uso del termine 'prescrizione' (applicato a medicine complete di indicazione terapeutica, norme estese alla preparazione e all'uso dei rimedi), da quello di 'ricetta' (applicato a formule assai semplificate, limitate all'indicazione dei componenti, attestate anche singolarmente su foglietti di papiro ed *ostraca*). Con 'prescrizione' e 'ricetta' identifico perciò tipologie leggermente differenti di testi. Definisco col termine 'ricettario' un testo poco elaborato formalmente, che raccoglie ricette o prescrizioni; con 'manuale terapeutico' intendo

---

67 On standardization in papyrological metadata see REGGIANI 2017, 74–8.

uno scritto in cui si riconosce un'organizzazione compositiva e formale più complessa, ora pro-dotto nell'ambito dell'insegnamento della disciplina, ora non diverso dai modelli di 'trattato' terapeutico" [Andorlini 1993, 469–70 n. 22].[68]

4.2 Introduction and commentary

The possibility to add a "front matter" and a "line-by-line commentary" is currently allowed by both *Papyri.info* and the DCLP, though it has been poorly exploited so far. Some documentary samples have been produced in the framework of the born-digital editions described by L. Berkes in this volume; for the DCLP side, see R. Ast and H. Essler *ibidem*. The DCGMP project utilizes systematically this feature to provide a general introduction to each text and to record the main textual features that cannot be encoded within the text for the moment, namely technical descriptions[69] (see below for future integration with the *Medicalia Online* lexical tool) and parallel passages in both other medical papyri and literature (see below for the intertextual layer).

An earlier way of inserting short comment strings (mostly providing information about re-editions of the texts) within the inline markup, through the `<note>` tag (Leiden+: `/* */`), is possible but definitely not exploited nor really recommended (the Leiden+ guidelines warn: "use sparingly"!)

4.3 Translation

Translations of the original text in multiple modern languages are currently supported in the existing databases. The DCGMP policy is to produce at least an English translation of each text, but when a scholarly translation in a different language does exist, the preference is granted to that one. As long as translation is a means of interpretation, the possibility to align the original text with its translation(s)[70] is worth being explored, for instance through the *Medicalia Online* lexical platform (see below).

4.4 Materiality

The physical appearance of the papyrus is of the utmost importance for the papyrologists, who are deeply interested in the material aspect of the fragments.[71] Size and colour are the first physical features that are indicated in a traditional edition, and

---

68 Bertonazzi 2018a, 48–51 (see also pp. 51 ff.).

69 In compliance with one of the original goals of the *Corpus dei Papiri Greci di Medicina*, i.e. the historical-scientific perspective described by Andorlini 1997a, 23.

70 On translation alignment cf. e.g. Véronis 2000.

71 See the remarks by R. Ast and H. Essler in this volume.

since they are not recorded in the metadata catalogues, they should be indicated in the introductory matter (see above) or in new metadata fields. A digital picture could compensate for this, but it is not available for all papyri (see below).

Material features of the writing support are encoded directly in the text itself according to the current standards. As to this point, there are some *notabilia* that must be stressed because they slightly differ from the traditional editorial practice. Line numbers, for example, are to be indicated for each line (contrarily to what happens in most of the printed editions) in a standardized way (number-dot-space); words that wrap between two lines are indicated with a hyphen after the dot of the second line number, not at the end of the first line. Both Leiden+ procedures may seem rather unconventional to 'traditional' papyrologists, but they are grounded on XML requirements: numbers are related to "line break" tags (`<lb/>`) that must open each new line of the encoded text; hyphens represent the attribute `break="no"` in the same `<lb/>` tag, meaning that the new line does not break the word.[72] In the HTML output things are brought back to the traditional display (line numbers grouped by five, hyphens at the word break).

Writing sides (recto/verso, folios in codices) and multiple fragments are encoded as document divisions (XML `<div type="textpart">`). This tag deploys an `n` attribute, which expresses the number/letter identifying the fragment/folio (or the letters `r/v` for recto/verso), and a `subtype` attribute, defining the type of part: `"fragment"`, `"folio"`, but also `"column"` or `"part"` if the text is divided into different layouts or sections even within the same writing side. By the way, this is a good way of dealing with texts that are composed by several sub-texts, like e.g. collections of letters or recipes.[73] Divs can be nested if needed, and each text block is anyway enclosed by an `<ab>` tag ("anonymous block"). In Leiden+, Divs are introduced by the tag `<D=` followed by the said attributes preceded by dot (e.g. `<D=.r` for recto, `<D=.1.fragment` for fragment 1), the text block by `<=`. Every text tag must be closed at the bottom, paying attention to the correct order (divs are opened before `<ab>` at the beginning, and symmetrically closed at the end).

The most remarkable physical feature of the papyri is fragmentation. This usually results in marginal breaks (printed as rows of dashes at the top and/or at the bottom, closed square bracket on the left, open square brackets on the right) and in-text gaps (represented as square brackets surrounding some indication of the missing text, which may or may not be supplemented). They are currently encoded as in-text markup; however, the digital concept of gap, according to the TEI/EpiDoc canons, is slightly different from the traditional one, and it deserves some comments. Each unsupplied break or lacuna is indeed treated as missing text, and all types of missing

---

72 See the contribution by G. Celano in this volume for the problems given by non-breaking lines in the digital papyrus texts.

73 An attempt of this can be found at http://litpap.info/dclp/60175 (P.Oxy. IX 1184, Hippocratic letters).

text are handled with the `<gap>` XML tag, which is used to encode both lost and illegible portions of text (the latter being usually printed – and are displayed – as series of dots). The `<gap>` attribute `reason` distinguishes the two cases (`"lost"` or `"illegible"`, plus `"ellipsis"` if the text is missing because left untranscribed by the editor), while the attribute `unit` specifies if we are dealing with just a number of characters or with entire lines. The extension of the missing text is defined by the attributes `extent` (`"unknown"` number of chartacters or lines), `quantity` (known number of characters or lines), `atLeast` / `atMost` (approximate range calculation). A `precision` attribute set to `"low"` indicates the uncertainty of an extension. Leiden+ syntax developed around the use of dot, after the print conventions of indicating illegible characters by means of dots: a dot followed by a number or a range (and by the indication `lin` when dealing with lines) indicates illegible text; the same, but preceded by the indication `lost`, marks lost text. If the dot is preceded by `vestig`, an element `<desc>vestiges</desc>` is added to the `<gap reason="illegible">` tag, in order to encode generic 'traces' (which is indeed the HTML output). Untranscribed text is marked differently (see below), as are supplied gaps, though from the papyrological viewpoint they are actually the same facts as the unsupplied ones (see below).

Unclear characters are another good example of how semantic markup differs from traditional print editions. In the latter, any unclear letter is marked with an underdot, either with the letter on its top (if legible) or not (if illegible). In the digital edition, illegible characters are non-textual portions marked with the `<gap reason="illegible">` tag described above, while unclear but legible characters are text portions marked with an `<unclear>` tag. Leiden+ utilizes the regular Unicode underdot in the latter case, while in the former the dot is recalled with a full stop followed by the number or range of unclear characters. In the HTML display, they become both underdots.

The close relationship between the text and its support is the core focus of the CRMtex project (http://www.cidoc-crm.org/crmtex), which provides tools for managing the study and publication of ancient handwritten documents and may be taken into consideration for developing new strategies in the digital edition of papyrus texts too.

4.5 Palaeography

Annotating palaeography is a huge task. Beside a general palaeographical description of the handwriting, which may well be detailed in the front matter, the possibility to mark up each single character is particularly tricky. Apart from its extreme intricacy, such a task should be preceded by a huge effort to standardize palaeographical terms and descriptions, which are notoriously idiosyncratic and inconsistent. Text

alignment with the digital picture can help: this is precisely the purpose of the *Anagnosis* project, conducted at Würzburg by Holger Essler, which may eventually come to the automatic recognition of the characters.[74]

Some visual characteristics of the written text are encoded with an appropriate markup that describes the appearance of lines, words or single characters with special display features. Lines that are written perpendicular or inverse with respect to the main body of the text can be encoded with a `rend="perpendicular"/"inverse"` attribute of the `<lb/>` tag (in Leiden+, this is obtained by putting the line number in brackets, typing a comma+space instead of dot+space, and then the appropriate attribute value). Similarly, ancient text highlights (taller characters, superscript, subscript, supraline, underline) are tagged with a `<hi>` element, with a `rend` attribute specifying the kind of highlighting (standard values: `"tall"`, `"superscript"`, `"subscript"`, `"supraline"`, `"supraline-underline"`). Leiden+ equivalents are shaped in a graphical appearance that hints to the text display on the papyrus (respectively: `~||x||~tall`; `|^x^|`; `\|x|/`; `‾x‾`; `=x=`). It must be noted that currently the use of this markup is deprecated when the highlighting describes an abbreviation (see below). A text written inside a box is encoded with a milestone element (`<milestone rend="box" unit="undefined"/>`; Leiden+: `###`).[75]

Another palaeographical feature that can be encoded with the current markup is the handshift (`<handShift new="m2"/>`; Leiden+: `$m2`; displayed as (hand 2)); for the use of this tag see also Marja Vierros' chapter in the present volume (which – by the way – contains also an interesting discussion about palaeographical metadata).

4.6 Text

At the core of the papyrus fragment, text as a linguistic fact deserves the highest and deepest attention, for both the peculiarities of the language of the papyri in general and the specific relevance of technical language in small corpora like the medical writings.[76] Digital annotation is a fundamental practice in the linguistic study of a corpus of texts:[77] it allows to describe, record, interpret and analyse linguistic information at several levels, in which each layer corresponds to a particular category of

---

74 See below and the *Anagnosis* section of the chapter by R. Ast and H. Essler in this volume; cf. Reggiani 2017, 151 ff.

75 Cf. Corazza 2018a; see below for 'milestones'.

76 The lexical and linguistic study has always been a primary purpose of the *Corpus dei Papiri Greci di Medicina*: cf. Andorlini 1997a, 24.

77 On the definition of linguistic corpus cf. Sinclair 1996; in general on corpus linguistics cf. Lüdeling – Kytö 2008–09; Lüdeling 2011; and see M. Vierros and I. Bonati in this volume. On the theoretical and practical correctness of treating Greek medical papyri as a proper textual corpus I think

relevant information.[78] Multiple levels of linguistic annotation of papyrological relevance can be outlined.[79]

4.6.1 Part-of-speech annotation

The basic annotation layer, related to the analysis of the parts of speech – also known as treebank 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 investigate formulaic uses and writing skills,[80] to find out influences or interpolations between authors, or the presence of literary echoes in technical or documentary texts.[81] The entire technical textual strategy deployed by medical authors[82] could be studied in this way. Analysing in depth and comprehending the syntactic structure of texts would allow also to solve problems of interpretation and attribution,[83] or even

---

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.

78 Cf. REGGIANI 2017, 178 ff., and M. Vierros in this volume.

79 Cf. REGGIANI 2015 and 2016a; BERTONAZZI 2018b.

80 Cf. MARAVELA – REGGIANI 2018. ROUED-CUNLIFFE 2014 described how digital encoding can prove useful for the analysis of grammar patterns of an ancient textual corpus (the Vindolanda tablets). A seminal project on annotating a corpus of private letters on papyrus, conducted by S. Porter and M. O'Donnell, has produced a number of valuable observations about modes and tenors of discourse, structures of information, semantic patterns, and so on (PORTER – O'DONNELL 2010).

81 "[L]a possibilità di identificare alcuni papiri con trattazioni di un autore tramandato solo indirettamente inserisce tasselli nuovi nella complessa stratificazione della trasmissione indiretta, soprattutto quando sono i papiri i soli testimoni diretti di autori tramandatici per *excerpta* e citazioni (Apollonius Mys, Heras, Heliodorus, Herodotus Medicus)" (ANDORLINI 1997a, 22).

82 ANDORLINI 2006 pinpointed the existence of an expressive strategy of medical technical texts: "L'osservazione di tali fenomeni, e del loro riproporsi costantemente nella tradizione dei testi medici greci su papiro, permette di riconoscere diverse fasi e livelli in cui il sapere tecnico contenuto nella ricetta medica veniva materialmente veicolato al lettore/consumatore attraverso moduli espresssivi e dispositivi tecnici, visivi, fisici, che formano una sorta di koinè, un tutt'uno tra lingua tecnica e scrittura speciale dei testi. Di qui la suggestione di rintracciare una specie di 'gergo' nei connotati di quel particolare linguaggio criptico, grafico ed espressivo, che comunica all'interno di una determinata categoria professionale: il medico, gli altri medici (i colleghi), il farmacista, il commerciante di farmaci, il paziente. Si tratta di modi speciali di usare parole e segni attraverso i quali le competenze medico-terapeutiche tendono a specializzarsi all'interno di una corporazione di addetti alla professione medica" (p. 153).

83 "L'analisi sintattica attraverso l'annotazione in un cosiddetto 'treebank' potrebbe mostrare più chiaramente la struttura del testo e facilitare il confronto tra il testo veicolato dal papiro e la tradizione

only to understand the exact meaning of a text (let us consider for instance the case of schematic prescriptions – e.g. P.Oxy. VIII 1088, http://litpap.info/dclp/63118 – 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, but papyrology too has made important progress, with the project *Sematia*, aimed at facilitating the linguistic tagging of the documentary papyri encoded in *Papyri.info* and described by Marja Vierros in this volume. Another possible way to linguistic annotation of the papyri is explored by Giuseppe Celano in the present book as well. The literary side has been unfolded by the *Grammatically Annotated Philodemus* project, conducted by Daniel Riaño Rufilanchas and Holger Essler (Würzburg) and aimed at deeply annotating the Greek philosophical papyri from Herculaneum on morphological, grammatical, semantic, stylistic layers.[84] Fragmentation is of course an issue when one decides to perform linguistic analysis: phrases, sentences, words are broken and it is not rarely difficult to understand the syntax, not to say to tokenize the words.[85] These are problems that digital tools must unavoidably face, and which an infrastructure based on multiple interconnected layers may feasibly overcome.

---

manoscritta, soprattutto nel caso di papiri per cui si sospetti una possibile paternità" (BERTONAZZI 2018a, 74). The case of surgical author Heliodorus is paradigmatic: "l'analisi del lessico tecnico dei papiri chirurgici ha portato a individuare paralleli testuali tra testo tramandato su papiro e tradizione manoscritta, talvolta significativamente stringenti come nel caso di P.Strasb. inv. 1187 e diversi passi di Eliodoro *ap*. Oribasio. Alcuni altri papiri (P.Lond.Lit. 166, P.Gen. inv. 111, P.Fuad.Univ. 1, P.Ryl. 3.529), come già notato dagli studiosi, sono caratterizzati da una forte presenza di 'lessico eliodoreo' e da alcune peculiarità proprie del *modus operandi* del chirurgo, come la predilezione di interventi chirurgici che siano il più sicuri possibili per il paziente, nonché del *modus scribendi*, come il ricorso frequente alla prima persona – singolare o plurale –, la definizione con esattezza delle posizioni 'topografiche' della parte operata (dentro, fuori, sopra, sotto), e una sostanziale semplicità delle strutture sintattiche usate. Ad oggi, i tentativi di attribuire i papiri citati alla paternità di Eliodoro si sono basati quasi esclusivamente su criteri lessicali nel confronto tra il testo tramandato su papiro e sui capitoli di Oribasio che portano la titolatura 'da Eliodoro'. Una nuova possibile strada offerta dalle nuove tecnologie della papirologia digitale è quella costituita dall'annotazione sintattica dei testi: un'analisi più accurata non solo del lessico, che come è noto è la parte più 'volatile' della lingua, ma delle strutture morfologiche e sintattiche dei passi del compilatore tardo in sinossi con i testi dei papiri, sia pure nella limitatezza delle pericopi testuali preservate, potrebbe gettare nuova luce anche su questo aspetto tra i più incerti quanto stimolanti della ricerca" (BERTONAZZI 2018a, 242–3). Marja Vierros has recently presented at the workshop "Act of the Scribe: Interfaces Between Scribal Work and Language Use" (Athens, April 6–8, 2017) some preliminary remarks on *Applying Modern Authorship Attribution Methods to Papyri and Ostraca* (abstract at http://blogs.helsinki.fi/actofscribe/workshop): cf. REGGIANI 2017, 185.

84 Cf. REGGIANI 2017, 181; R. Ast and H. Essler in this volume.

85 Cf. RIAÑO RUFILANCHAS 2014, 160–1; ESSLER – RIAÑO RUFILANCHAS 2016, 498; and the observations by R. Ast and H. Essler, M. Vierros, and G. Celano in the present volume.
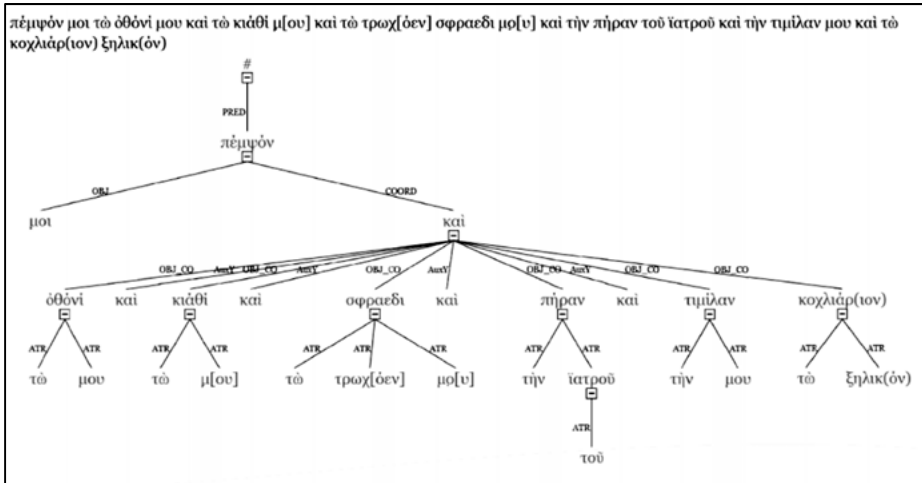
πέμψόν μοι τὼ ὀθόνί μου καὶ τὼ κιάθι μ[ου] καὶ τὼ τρωχ[όεν] σφραεδι μο[υ] καὶ τὴν πῆραν τοῦ ἰατροῦ καὶ τὴν τιμίλαν μου καὶ τὼ κοχλιάρ(ιον) ξηλικ(όν)

Fig. 3: Sample treebanking of GMP II 10, medical letter (from Reggiani 2015).

4.6.2 Lemmatization

An annotation layer of lemmatization, 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, which has always been a relevant research focus of Isabella Andorlini's concept of the medical corpus.[86] Such a sort of layer would represent an important bridge to connect the textual database to the related project *Medicalia Online*, consisting in an extensive lexical reference platform for ancient medical technical terms, as described by Isabella Bonati in this volume.[87] Systematic links to the lexical records (and the other way around) could contribute to create a dynamic lexicon[88] of medical technical terms in the Greek papyri. In addition, as Joanne Stolk observes in this volume, the possible deployment of a lemmatization layer would help encoding linguistic variation more properly, while encoding lexical information would be helpful for the creation of word reference indices.

---

86 Cf. Andorlini 1997a, 24. For more recent works on this topic see Bonati 2016a, 2017, 2018a, 2018c, and Bertonazzi 2018a. For a parallel exploitation of digital encoding for the development of vocabulary analysis, cf. Roued-Cunliffe 2014 apropos of the Vindolanda corpus.

87 Cf. also Bonati 2018b and 2018d. On the connection between digital editions and *Medicalia Online* cf. also Bertonazzi 2018a, 43–8 and 73–4, and 2018b.

88 On "the interdependence of lexica and new editions" cf. Essler – Riaño Rufilanchas 2016, 492. Roued-Cunliffe 2014 speaks of "integrated indexing".

4.6.3 Abbreviations

Abbreviations are another striking point. Medical writings (prescriptions above all, but not only) make a particularly extensive use of abbreviated words,[89] developing a proper "graphical-expressive jargon";[90] given their technical nature, it would be extremely useful to investigate their use, e.g. whether there is any underlying pattern. As to now, abbreviations are to be encoded in the same way as the documentary papyri, that is according to the type of expansion – resolved or unresolved, distinguished on the ground of the XML syntax. Resolved abbreviations are encoded as "expansions", with the `<expan>` tag enclosing the text spelled out and the `<ex>` tag enclosing the text abbreviated (Leiden+: double set of brackets, one enclosing the whole word and the other one enclosing the expanded abbreviation); unresolved abbreviations are encoded as "abbreviations", enclosed by the `<abbr>` tag (Leiden+: `(|x|)`). Any attempt to encode the type of abbreviation (e.g. by raised letter or by overline) is currently deprecated.[91] I strongly hope that in the future this level of annotation may be taken into consideration, since abbreviating strategies are relevant for the correct transcription and interpretation of texts, as in P.Strasb. inv. 1187 (http://litpap.info/dclp/59968), which

> exhibits two cases of allegedly abbreviated words that have been object of interpretative discussion. At ll. 11 and 14 two ν overlined with a horizontal stroke (belonging to a plural genitive and a nominative respectively: -ω⁻) are clearly legible; these strokes are abbreviation marks according to FAUSTI 1989, 158, *contra* MARGANNE 1998, 68, following *ed.pr.* for the latter, which supplies the ν as omitted by the scribe, in angle brackets. The presence of the overline strongly suggests that we are indeed dealing with abbreviated words: therefore, though relying by rule on the more recent edition, [for the digital edition] it has been chosen to follow the *editio altera*, marking the abbreviations according to the current Leiden+ conventions, though preserving the reading of the *editio tertia* in an |ed| tag.[92]

In the described case, a correct understanding of the abbreviation mark proves essential in the text editing and encoding. Moreover, special ways of expressing combinations of characters or even entire words cannot be encoded but in the standard, simplified way: for example, to limit ourselves to the cases of P.Ant. III 127 (http://litpap.info/dclp/65340) mentioned by CORAZZA 2018b, the sinusoid for αι and the peculiar sign ·//· for εισι, which must be encoded as whichever symbol `<expan><ex>εισι</ex></expan>` (Leiden+: `((εισι))`), losing interesting pieces of information.

---

[89] Cf. e.g. the case of the Antinoupolis papyri described in CORAZZA 2018b.

[90] Cf. ANDORLINI 2006.

[91] Cf. REGGIANI 2018b, and see L. Berkes in this volume. As he notes, text-image alignment could be a good compromise: but searching for the different abbreviation types would not be possible as well.

[92] BERTONAZZI 2018b; cf. BERTONAZZI 2018a, 67.

A tentative proposal, based on a preliminary survey conducted on the different abbreviation typologies in the Greek medical papyri and on the TEI/EpiDoc XML guidelines, envisions the following possible instances (apart from 'traditional', 'simple' abbreviations):[93]

– <u>Supralinear abbreviations</u> (e.g. χαλκάν^θ(ου): PSI X 1180a iii,12). The superscripted letter may be tagged as any normal superscripted letter (`<hi rend="superscript">` tag); such a combination is already possible and correct in the current Leiden+ syntax, yet deprecated by the official guidelines. XML: `<expan>χαλκάν<hi rend="superscript">θ</hi><ex>ου</ex></expan>`; L+: `(χαλκάν|^θ^|(ου))`.

– <u>Abbreviations by stroke</u> (<u>horizontal</u>: e.g. τῶ̄(ν), P.Mich. XVII 758 H verso,2; <u>vertical</u>: e.g. ξ|(ηρόν), P.Mich. XVII 758 H verso,3; <u>slanting</u>: e.g. χαλβάν/(η), P.Mich. 758 H,11; <u>sinusoid</u>: e.g. γίγνετʃ(αι), P.Ant. III 127, i b, 6). The strokes may be encoded through the EpiDoc `<am>` tag ("abbreviation mark")[94] and further defined as non-alphabetic glyphs (see below) as follows: `<expan> <abbr> τῶ <am> <g type="horizontal-stroke"/> </am> </abbr> <ex>ν</ex></expan>` = `(τῶ *horizontal-stroke*(ν))`; `<expan><abbr> ξ<am><g type="vertical-stroke"/></am></abbr><ex>ηρόν</ex></expan>` = `(ξ *vertical-stroke* (ηρόν))`; `<expan> <abbr> χαλβάν <am> <g type="slanting-stroke"/> </am> </abbr> <ex>η</ex></expan>` = `(χαλβάν*slanting-stroke*(η))`; `<expan><abbr>γίγνετ<am><g type="sinusoid"/></am></abbr><ex>αι</ex></expan>` = `(γίγνετ *sinusoid*(αι))`. Note that such combinations are correct in the current Leiden+ syntax, but the strokes need to be rendered properly in the HTML output; moreover, the `<am>` tag is not supported by the platform.

– <u>Discontinuous abbreviations</u> (e.g. μ(ε)τ(ά): MPER n.s. XIII 9, 1). This type of abbreviation is already normally working in the SoSOL environment. `<expan>μ<ex>ε</ex>τ<ex>ά</ex></expan>` = `(μ(ε)τ(ά))`.

– <u>Abbreviations by monogram</u> (e.g. σχι(στοῦ); πρ(ός); χρ(ῷ)).[95] This type exploits the way in which monograms are marked up in EpiDoc and Leiden+,[96] i.e. a 'g-type' with indication of the letters that are interwoven to form the monogram. `<expan> <abbr> <am> <g type="monogram">σχι</g> </am> </abbr>`

---

93 In general, on abbreviations in papyri see e.g. CLARYSSE 1990, DEGNI 1999, and GONIS 2009; with special regards to documentary texts, BELL 1953 and BLANCHARD 1974; for literary papyri, MCNAMEE 1981 and 1985. A typological work on the abbreviations in medical papyri has been preliminarily conducted by L. Iori and M. Centenari in the framework of the *Corpus of the Greek Medical Papyri Online* project (cf. http://www.papirologia.unipr.it/eventi/GDS/2010/centenari-iori.html).

94 Cf. http://www.stoa.org/epidoc/gl/latest/trans-abbrevmark.html.

95 On the relevance of the monogram χρ(ῷ) see ANDORLINI 2018.

96 Cf. http://papyri.info/docs/leiden_plus *s.v.* "Non-alphabetical character with symbol".

```
<ex>σχιστοῦ</ex></expan>= ((*monogram,σχι*σχιστοῦ)) ;<expan>
<abbr> <am> <g type="monogram"> πρ </g> </am> </abbr> <ex>
πρός </ex> </expan> = ((*monogram,πρ*πρός)) ;<expan> <abbr>
<am> <g type="monogram"> χρ </g> </am> </abbr> <ex> χρῶ
</ex> </expan> = ((*monogram,χρ*χρῶ)).
```
This may apply to some symbols for units of measure too, e.g. λί(τρα), ο(ὐ)γ(χία), etc., which may make easier a systematic study of quantities and dosages in the ingredient use. Number digits and values can be easily encoded in the current way (XML: `<num value="16">ις</num>`; Leiden+: `<#ις=16#>`).

4.6.4 Linguistic variation

The topic of linguistic variation is the most intriguing and difficult to handle. As noted above, traditional critical editions tend to overcome any fluctuation in favour of a reconstructed text, while fluctuations are actually fundamental for the phenomenology of the written text. Linguistic variation in the papyri has already been extensively investigated by Joanne Stolk, who resumes her thoughts from the digital perspective in this same volume. I would like just to focus on some relevant points, to introduce the problem of linguistic variation in the medical papyri. The current markup of what I call textual fluctuations – handled by the `<choice>` tag, "indicating that [the readings] are two editorial versions of the same span of text, and should be read as alternatives, not shown side by side"[97] – distinguishes between "corrections" of outright, well recognizable scribal mistakes (`<corr>` tag marking the correction, `<sic>` tag marking the original reading; Leiden+: `<:`*correction*`|corr|`*original*`:>`) and "regularizations" of phonetic misspellings (`<reg>` tag marking the regularization, `<orig>` tag marking the original reading; Leiden+: `<:`*regularization*`|reg|`*original*`:>`).[98] Though the treatment of 'regularizations' has been improved during the history of the

---

97 http://www.stoa.org/epidoc/gl/latest/trans-regularization.html: "It is most common to mark a regularization of this kind at the level of the whole word, rather than of individual characters affected [...]. This will make it easier to generate an apparatus reading for the regularized form (or the original form, depending on which you want to privilege), but it may also be impossible to identify individual affected characters in a dialect spelling or grammatical form. On the other hand, tagging the individual characters might make it easier to index or search for specific features, such as the iotacism of ι and ει". See J. Stolk's observation (in this volume) that apparatus regularizations/corrections work as textual equivalents and not 'better' substitutes of the original text.
98 For some case studies cf. BERTONAZZI 2018a, 63–4.

papyrological databases, by moving the display of the original spelling from the apparatus to the main text and vice versa,[99] showing a stronger care for the phenomenology of the papyrus text, and though the meaning of the `<choice>` tag points to alternative encodings of the same text portion, the fact is that we are still dealing with a differentiation between a form that is considered as 'standard' or 'regular' and a form that deviates from it. Phonetic fluctuations like σμύρνη / ζμύρνη in the medical papyri[100] (but see also some relevant cases in the documentary papyri, like χύτρα / κύθρα[101] and ἔ(ι)σοπτρον / ὄσυπτρον[102]) show that not always is it easy to define what is the 'conventional' spelling and what is the 'deviation', so that a layer capable to align the 'variants' to each other, word by word, rather than categorizing them in a sort of hierarchy, may be much welcome.[103]

Diachronic and synchronic fluctuations – depending on the evolutions and transformations of Hellenistic Greek language and on the rise of personal or geographical substandards[104] – do occur in the medical papyri, but their existence not rarely points

---

99 In the earlier *Duke Databank of Documentary Papyri*, "the conventional *koine* form is given first, followed by numbered braces enclosing the scribe's form or the edition's misprint: e.g., ὄνομα {4ωνομα}4 shows that the scribe has misspelled ὄνομα, ὑπὲρ {5υπαρ}5 that he wrote epsilon over alpha, αὐτοῦ {6αυτω}6 that he miswrote dative for genitive, Ἀθὺρ {7Ἀθὺς}7 that the edition has a misprint for Ἀθὺρ" (Willis 1984, 169–70); cf. Reggiani 2017, 216. This slight prominence given to the 'standard' form was retained in the first stages of *Papyri.info*, where it was included in the main digital text, whereas the original form as written by the scribe – marked with |reg|, |corr| or the former |orth| tag (cf. Reggiani 2017, 236 n. 119) – was displayed in the apparatus. As of September 2011, the two elements in the |reg| tag have been swapped with each other (cf. http://digitalpapyrology.blogspot.it/2011/09/just-posted-to-papylist-dear-colleagues.html). This required a huge effort, because the ancient reading was originally transcribed diplomatically without spirits and accents, but its inclusion in the text made it necessary to add them (cf. Reggiani 2017, 224).
100 "Nei papiri è scritto quasi regolarmente ζμ-" (Andorlini 1981, 61 n. 54), which conversely should be a 'deviating' spelling of 'regular' σμύρνη (cf. Gignac 1976, 121–2).
101 Cf. Bonati 2015.
102 On this peculiar double fluctuation cf. Bonati – Reggiani 2018.
103 Cf. e.g. Boschetti 2007 apropos of philological variant alignment; further discussion in Reggiani 2018a. The current platform also allows for handling language shifts, i.e. the markup of a language or script different than the main document's default. This is rendered with the tag `<foreign>` and the `xml:lang` attribute, the value of which may be `grc` for Greek words in a Latin text, `la` for Latin words in a Greek text, `grc-Latn` for Greek words in Latin characters, `la-Grek` for Latin words in Greek characters, and so on. In Leiden+ it is marked `~|x|~grc` and the like. If characters or lines in a different language or script are omitted by the editor, this is indicated with the `<gap reason="ellipsis">` tag (see above) including a `<desc>` tag filled with the appropriate language (e.g. Coptic, Demotic; Leiden+: `(Lang: Coptic 1 line)` etc.). It is also possible to mark up cross-language equivalencies, for example giving the Greek correspondent of a Coptic term. For this task, the current system exploits the 'regularization' tag by adding an `xml:lang` attribute, e.g. `<choice> <reg xml:lang="grc"> ἀρακος </reg> <orig> ⲁⲣⲁⲕ </orig> </choice>` (Leiden+: `<:ἀρακος=grc|reg|ⲁⲣⲁⲕ:>`). The explanatory note goes into the apparatus accordingly.
104 Cf. Reggiani 2018a and 2018c.

to deeper levels of textuality, with reference to ancient literacy and intertextual relations (see below), and therefore deserve a very peculiar attention. For example, in P.Aberd. 124, i = GMP I 1, i (II cent. AD, http://litpap.info/dclp/63334), a papyrus fragment preserving chapter 37 of Hippocrates' treatise *De fracturis*, at l. 14, where all the codices (and the editions) have the 'regular' Ionic form πήχεος, the papyrus shows clearly π]ήχεως, the *Koine* form, which looks like an 'interference' of a typical 'linguistic variation' pertaining to the language of the documentary papyri, where it would be the standard form.[105] Perhaps even more significant are the following cases. P.Oslo inv. 1576, a fragment of a catechism dealing with tumour-like diseases,[106] partly overlaps with the text of P.Oxy. LXXX 5239 (both II–III cent. AD). The latter is more likely a 'treatise' than a questionnaire, as its editor David Leith notes (see below for this distinction), and the difference may be perceived from the lack of *eistheseis* in its questions. The scarceness of the surviving portions of text makes it hard to say whether the questionnaire derives from the treatise or they are two different outcomes of a same ascendant (see below for intertextual relations). As far as the extant parallel text is concerned, the wordings diverge from each other only for one variant: ὑδροκήλη (P.Oslo, l. 5) vs [ὑ]γροκήλη (P.Oxy., l. 15). The latter is usually considered as a minority variant (LSJ, quoting Poll. IV 203) of the former, used e.g. by Ps.Gal. *Def.med.* 424 = XIX 447,12–13 K., but it is in fact attested three times among the medical writers.[107] Are we facing a trivialization in the Oslo papyrus, or a simple phonetic variant in the Oxyrhynchus papyrus, or just two different traditions bearing the same degree of 'correctness', attesting to a fluid notion of technical language? Moreover, in the following line of the Oslo papyrus (not paralleled by its Oxyrhynchus counterpart any more) we read ἐρυτρ[οειδῆ, which looks like a phonetic variant of ἐλυτροειδής "lid-like", "cover-like" (attribute of one of the membranes enveloping the scrotum). *Rho* for *lambda* is indeed a very frequent phonetic exchange in the language of the Greek papyri,[108] but the same variation is to be found among the manuscripts preserving Ps.Galen's *Introductio seu Medicus*, containing a descriptive passage (XIV 719,5– 10 K.) of the same anatomical part.[109] A similar case is offered by P.Coll.Youtie I 4v

---

105 A comparable case is τοῖς (and the following forms supplied accordingly) in P.Fay. 204,9 (http://litpap.info/dclp/60181) vs Ionic τοῖσι of the rest of the tradition of Hippocratic aphorisms.

106 MARAVELA – LEITH 2007. The papyrus will be republished in the forthcoming third volume of the *Papyri Osloenses*. I am most grateful to Anastasia Maravela for sharing her drafts of the new edition and for discussing with me some textual and linguistic details.

107 Orib. *Syn.Eust.* III 28, 6 and 9 = CMG VI 3, p. 75, 15–16 and 21 Raeder; Steph. *In Hp. Progn.* II 1 = CMG XI 1,2, p. 140,25 Duffy. The case resembles – *mutatis mutandis* – that of ὄσυπτρον, the above-mentioned 'deviant' form of ἔ(ι)σοπτρον "mirror", for which BONATI 2016a, 246–8 already proposed the rank of substandard.

108 Cf. GIGNAC 1976, 105.

109 The previous editors corrected it in ἐρυθροειδοῦς, but the newest Belles Lettres edition (PETIT 2009) prints ἐλυτροειδοῦς (XII 11, p. 40,1; see PETIT 2009, xcvi–xcix for the description of the manuscript tradition). Quite interestingly, the author of the treatise came possibly from Egypt (cf. PETIT

(http://www.litpap.info/dclp/64118), a collection of prescriptions dated around the III cent. AD, in which φλοῦς at l. 8 ("reed bed") could be a spelling variant of φλοιός ("bark") just as φλοιόν in Dsc. III 147 has a variant reading φλοῦν.[110] In a completely different type of text, P.Oxy. XIX 2221r + P.Köln V 206r (http://litpap.info/dclp/61916), a I-century AD commentary to Nicander's *Theriaka* attesting interesting textual variants (see below), at l. ii,29 we read βορεῖται vs βοτεῖται codd. (Nic. *Ther.* 394), which looks like the genuine form; the reading of the papyrus is a phonetic variant of the φορεῖται to be found in the ancient *scholia* to that passage.[111] Once more time, the impression is that we are facing a peculiar intersection of multiple literacies, emerging at the phonetic level but implying deeper meanings that cannot be flattened in a traditional apparatus.

Even seemingly outright syntactic 'mistakes', in such a technical corpus as the medical papyri, can conceal deeper levels of meaning: an established prescriptive formula like ὕδωρ χρῶ "use with water" goes far beyond the apparent anacoluthon (ἐν ὕδατι would be expected), becoming a distinctive mark of medical recipes, and must be treated accordingly.[112]

A very tentative example of annotation of a medical papyrus: GMP II 10.

ID = *Identifier*; **T** = *Tokenization*; **PoS** = *Part of Speech*; **H** = *Syntactic head*; **R** = *Syntactic relation* [H + R + ID = *Syntactical annotation*]; **N** = *Normalization*; **E** = *Kinds of errors* [VE = vowel exchange; BD = broken diminutive; ML = missing letter; EC = error corrected; CE = consonant exchange]; **L** = *Lemmatization*; **TT** = *Technical terms*; **A** = *Abbreviations*; **M** = *(Dia)critical marks and punctuation* [The other levels of annotation do not apply to this instance. Note that PoS and syntactical taggings are simplified for reason of space]. I am very grateful to Dr. Giuseppe G.A. Celano for his essential help in the grammatical/linguistic annotation.

ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14
T	πέμψόν	μοι	τό	ὀθόνι	μου	καὶ	τό	κιαθί	μ[ου]	καὶ	τό	τρωγ[όεν]	σφραξεῖ	μο[υ]
PoS	verb	pron.	art.	noun	pron.	conj.	art.	noun	pron.	conj.	art.	adj.	noun	pron.
H	0	24	4	24	4	24	8	24	8	24	13	13	24	13
R	PRED	ATR	ATR	OBJ_CO	ATR	AuxY	ATR	OBJ_CO	ATR	AuxY	ATR	ATR	OBJ_CO	ATR
N			τό	ὀθόνιόν			τό	κυάθιον			τό		σφραγίδιόν	
E		VE	BD		VE		VE+BD		VE			AS+VE+ BD		
L				ὀθόνιον				κύαθος					σφραγίς	
TT				instrument				container					instrument	
A														
M														

ID	15	16	17	18	19	20	21	22	23	24	25	26	27
T	καὶ	τήν	πήραν	τό⁵	[.]ιατροῦ⁵	καὶ	τήν	τιμιλίαν	μου	καὶ	τό	[χ]κοχλιάρ/(ιον)	ξηλικ/(όν)
PoS	conj.	art.	noun	art.	noun	conj.	art.	noun	pron.	conj.	art.	noun	adj.
H	24	17	24	19	17	24	22	24	22	1	26	24	26
R	AuxY	ATR	OBJ_CO	ATR	ATR	AuxY	ATR	OBJ_CO	ATR	AuxY	ATR	OBJ_CO	ATR
N				τοῦ	ιατροῦ			σμιλην			τό	κοχλιάριον	ξυλικόν
E					EC(?)			CE+ML+VE			VE	EC(CE)	VE
L			πήρα		ιατρός			σμίλη				κοχλιάριον	
TT			instrument		gen. term			surgery				container	
A				superscr.	superscr.							slanting stroke	slant stroke
M					diaeresis								

Fig. 4: a former attempt to outline some annotation layers for Greek medical papyri (Reggiani 2015).

2009, 1—li), which suggests that the phonetic variation could have worked both ways. I discuss this and the preceding case in Reggiani 2018a and 2018e.

110 Cf. T.T. Renner, *ad loc.*

111 The same phonetic exchange β/φ occurs elsewhere in the papyrus: see ἀμφίσφαινα for ἀμφίσβαινα in ii,9, 14, 15.

112 Cf. Andorlini 2006, 163, and 2018; Reggiani 2018a.

4.6.5 Transtextuality

As we saw, quite often linguistic phenomena may be clues to broader cultural facts. The complexity of textual phenomena in the Greek medical papyri (but not only!) can be effectively described through the concept of transtextuality as investigated by Gérard Genette since the Eighties. Transtextuality defines all the various possible relationships among texts ("all that sets the text in relationship, whether obvious or concealed, with other texts")[113] and encompasses several subcategories,[114] on which I will base the description of the next layers.

4.7 Paratext

Paratextuality is defined as the relation between one text and what surrounds the main body of the text: in Genette's theory, paratext is mainly composed of titles and headings, but we may add any other graphical device that comes along the text itself, including punctuation, which is not a common feature in papyrus texts and therefore deserves special treatment.[115]

In a writing system based on *scriptio continua* (i.e. not separating letters into words), punctuation is a way of facilitating the reading by separating words or groups of words. Single, double, triple dots occur irregularly with this function; in TEI/EpiDoc they are encoded as "non-alphabetical glyphs" with the tag `<g>` and the attribute `type` defining their nature (e.g. `<g type="middot"/>`, `<g type="dipunct"/>`, `<g type="tripunct"/>`). In Leiden+, the so-called 'g-types' are encoded by typing the attribute name between two asterisks. This is usually how other lectional signs (*apostrophe*, *diastole*, *stigmai*) and all graphical devices (check marks, deletion marks, parentheses, line fillers, strokes) work in this markup: a full list of what is supported by *Papyri.info* can be found at http://147.142.225.252/paptrac/wiki/gtypes, while new signs are being developed specifically for the DCLP.[116]

A small set of other signs, separating not letters or words but entire text sections, is encoded as 'milestones':[117] this is the case with *paragraphos* (`<milestone`

---

113 GENETTE 1992, 83, then GENETTE 1997, 1.

114 Cf. GENETTE 1992, 83–4, later developed in GENETTE 1997, 1–7.

115 On punctuation in the papyri cf. the overviews by TURNER 1987, 7–10, and CRIBIORE 1996, 81–3. For specific issues cf. DEL MASTRO 2017 (Herculaneum papyri) and FUNARI 2017 (historical fragments). For the particular care for paratext in the digital editions of literary papyri see the notes by R. Ast and H. Essler in this volume.

116 On ancient punctuation and encoding/annotating issues see the article by G. Celano in this volume. On filling marks in the papyri cf. BARBIS LUPI 1992. On diacritical and lectional signs see now also NODAR DOMÍNGUEZ 2017 and MCNAMEE 2017.

117 Cf. http://www.stoa.org/epidoc/gl/latest/trans-textpart.html for the difference between Divs (structural text parts) and milestones (non-structural text parts). A possible issue is that in fact e.g.

rend="paragraphos" unit="undefined"/>`), which has been supported by *Papyri.info* since the beginnings, and with the new additions developed for DCLP, i.e. *koronis* (`<milestone rend="coronis" unit="undefined"/>`) and forked *paragraphos* a.k.a. *diple obelismene* (`<milestone rend="diple-obelismene" unit="undefined"/>`).[118] They are all placed between any two lines of text and in Leiden+ are encoded as rows of four typographical characters depending on the sign (four hyphens for *paragraphos*: `----`, four equals for *koronis*: `====`; four combinations of dash + angle bracket for *diple obelismene*: `->->->->`).

The *diple* has been categorized as 'milestone' as well (`<milestone rend="diple" unit="undefined"/>`, Leiden+: `>>>>`), though this may create some issues, since *diplai* are frequently used in the margins to highlight a specific section, phrase or word, and encoding them as milestones would not be semantically correct.[119] A similar issue may arise when the *paragraphos* is used between two lines to separate two sections of a text but the logical division occurs within the preceding or the following line, as in PSI VI 718 = SB XXVI 16458 (IV AD, http://litpap.info/dclp/64564). This sheet, likely cut off from a small parchment notebook, contains part of a collection of prescriptions separated from each other by inline filling marks and interlinear *paragraphoi*. The last recipe starts in l. 12, following the end of the preceding one and after a separator mark, though the *paragraphos* is traced between ll. 12 and 13.
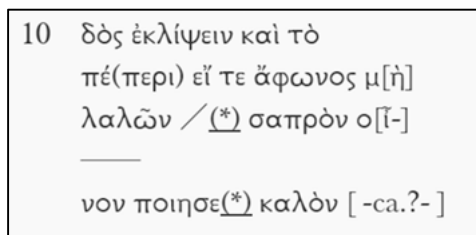
Fig. 5: SB XXVI 16458,10–13

```
<lb n="12"/>λαλῶν <g type="check"/></lem><rdg><choice><corr><expan>
πέ<ex>π</ex>ε<ex>ρι</ex></expan>ἰτέα<supplied    reason="omitted">ς
</supplied> φλοιός</corr><sic>πεειτεαφωνος</sic></choice> μ<supplied
reason="lost">α</supplied><lb n="12" break="no"/>λαχῶν <expan><ex>
```

---

paragraphoi may actually mark a subdivision between structural text parts, but the underlying rationale in the papyrological markup seems to be that they represent graphical separators or turning points marked by the ancient scribe.

118 On these peculiar signs cf. Barbis Lupi 1994 (*paragraphos*), Barbis Lupi 1988 (*diple obelismene*), Schironi 2010, 16–18 and passim (*koronis*).

119 A `<hi>` tag would probably be better: see below the case of *eisthesis/ekthesis*.

```
δραχμὰς</ex></expan>  <num  value="6">ϛ</num>.</rdg></app>  σαπρὸν  ο
<supplied reason="lost">ἱ̃</supplied>

<milestone rend="paragraphos" unit="undefined"/>

<lb  n="13"  break="no"/>νον  <choice><reg>ποιῆσαι</reg><orig>ποιησε
</orig></choice>  καλὸν  <gap  reason="lost"  extent="unknown"  unit=
"character"/>
```

In such cases, the sign is reproduced as in the original text, but the semantics is odd; moreover, word breaks between two lines separated by a *paragraphos* seem not to be handled by the searching engine. It is quite clear that rigid textual units lose relevance when one deals especially with technical texts, and a separate layer to record the paratext in all its multifarious relations with the text may prove useful.[120]

Blank spaces are a particular category of paratextual devices that deserves a thorough reflection. If the main purpose of punctuation is to divide text portions, then it is possible to think that any "space deliberately left blank [inside a text] is also to be considered as a mode of punctuation".[121] The traditional way of referring to deliberate blank spaces is the *vacat*, which is rendered with a `<space>` tag in TEI/EpiDoc (with the very same attributes as the `<gap>` mentioned above). Of course, encoding recurring blank spaces like those deployed in P.Col. IV 122 (official letter, 181 BC, http://papyri.info/ddbdp/p.col;4;122) to separate almost every word from one another would be impossible, if not in a different layer than text. Normally, the *vacat* is to be marked up when it introduces a significant break in the text.[122] Peculiar cases as in P.Oslo III 72,9 (medical treatise about epilepsy, II AD, http://litpap.info/dclp/63583), where (according to the editors' interpretation) the ancient scribe left a blank gap to pinpoint a controversial point, should be further (or differently) annotated in order to preserve the original intentions.[123]

The handling of blanks is connected to the problem of how to encode *ekthesis* and *eisthesis*, i.e. extension and indention of a line with the purpose of highlight particular phrases, which has never been taken into consideration before for documentary papyri. Among the medical texts, this device is frequently deployed in the questionnaires or catechisms. Such a text typology provides medical notions in a dialogue format, where a question about theoretical definitions or practical procedures is followed by

---

120 Diacritics and lectional signs added by different hands are another case of uneasy elaboration.

121 TURNER 1987, 8; cf. also CRIBIORE 1996, 83 ("Blank spaces can be used as punctuation").

122 *Vacat* can be used also to render columnation in particular layouts (lists, accounts, etc.), but the use is not standardized. See the chapter by L. Berkes in this volume for some remarks on the markup of layout in documentary papyri. Very recently, DICKEY 2017 has dealt with particular layouts of bilingual texts, where the columns are handled with blank spaces.

123 This case is currently encoded as an editorial apparatus note displaying the omitted text.

a more or less detailed answer.[124] Its use as a handbook, a reference tool for the doctors' preparation, is clear also from the complex set of devices employed to highlight the articulation of the text: questions are very often indented in *eisthesis*, and further marked with *paragraphoi*, line fillers or other lectional marks that introduce the answers as well. This *mise en page* reflects the central role played by the question-and-answer structure of the didactical tool,[125] and must be preserved when the texts are moved to any modern format. This is not only a matter of reproduction. In the overall framework of difficulty of recognizing textual genres[126] due to the fragmentary state of the scattered sources preserved to us, scholarship relies on any possible feature for a better understanding of ancient texts, and some very fragmentary texts have been identified as questionnaires just on the basis of the presence of blank spaces (P.Oxford Sackler s.n., II century BC;[127] more recently GMP I 6 and P.Strasb. inv. 849):[128] it is therefore unconceivable to encode such texts without paying attention to their paratextual garment.

It is tempting, at a first stage, to equate an *eisthesis* to an initial *vacat* and therefore to encode it like that. However, as we have to encode not the visual appearance of the text but its semantic core, we must be aware of the fact that we are not describing a certain extent of space intentionally left without characters, but a displacement of the line beginning to stress its relevance.[129] Its specular counterpart, *ekthesis*, makes the picture clearer: by no means can it be indicated by creating weird virtual *vacats* at the beginning of the surrounding lines. The current solution is to mark it as an attribute of the line: `<lb n="1" rend="indent"/>`, which in Leiden+ appears as `(1, indent)` – the same way marginal annotations are tagged (see below; for *ekthesis* the value `"outdent"` is to be used).[130] This seems to work fine, and is now fully supported by the DCLP platform also in terms of visual display.

124 Cf. Reggiani 2016b with earlier bibliography; also Bonati 2018e.

125 Cf. Andorlini 1999; Reggiani 2018h.

126 Cf. Andorlini 1997b, 159, and see above.

127 Cf. Barns 1949, 4–5. Online: http://litpap.info/dclp/65633.

128 Cf. Hanson – Mattern 2001, 72 and Magdelaine 2004, 63, respectively. Online: http://litpap.info/dclp/69007 and http://litpap.info/dclp/69028.

129 On the ecdotic relevance of line displacement in the system of the margins of the Greek literary papyri see Savignago 2008 (cf. also Turner 1987, 8).

130 In medical papyri *ekthesis* is somehow less frequent than *eisthesis*; a significant case is presented by Corazza 2018a. See also P.Oxy. XIX 2221r + P.Köln V 206r (http://litpap.info/dclp/61916), the aforementioned commentary to Nicander's *Theriaka*, where the lemmas containing the commented passages are highlighted by *ekthesis* (cf. Andorlini 2000, 39) and the comments are introduced by larger blank spaces, which might be considered as *eistheseis*.

| (Fig. 8a: visualizzazione del testo nel PE) | (Fig. 8b: visualizzazione del testo in DCLP) |

Fig. 6: A nice comparison between the earlier SoSOL preview display and the current DCLP rendering of *eisthesis* in P.Gen. inv. 111, catechism, http://litpap.info/dclp/63819 (Bertonazzi 2018a, 42).

However, a further problem arises if we consider that in some catechisms the questions do not start in a new line, but on the same line as the end of the previous answers, after a blank space that cannot be considered as a *vacat* for the same reasons as above. In this case, if we tagged the entire line as in *eisthesis* we would not represent the situation correctly, since the first part of the line is not really indented. A new solution might be to tag the question phrase with the TEI/EpiDoc XML <hi> element, which is used to sign "highlighted characters or words", "with a rend attribute specifying the kind of highlighting",[131] In our case, the value of the rend attribute would be "eisthesis", i.e. <hi rend="eisthesis"></hi> (or "ekthesis" in the other case), which is not supported by SoSOL currently.[132]

---

131 Cf. http://www.stoa.org/epidoc/gl/latest/trans-charactershighlighted.html.

132 Cf. Reggiani 2018h, where I advanced a further distinction of *eistheseis* according to their appearance in the texts. It is worth noting that encoding *eisthesis/ekthesis* as a <hi> element would prove helpful also when handling whole indented/outdented paragraphs (see the sample in the Conclusions below), instead of marking each line.

The `<hi>` tag handles also the markup of ancient diacritical signs originally added by the scribe,[133] which has always been well developed since the earlier times of *Papyri.info*. The canonical cases are accents (acute, circumflex), spirits (lenis, asper), diaeresis, either alone or in combination, marked with a `rend` attribute, the value of which corresponds to the name of the sign. Leiden+ markup is rather intricate (they must be added in the proper Unicode character inside a pair of brackets just after the appropriate letter, which in turn must be always preceded by an extra space, in whichever position it occurs in the word), but fortunately the editorial platform offers quite a helpful menu to automatically perform the task. It is worth noting that the presence of ancient diacriticals is noted in the apparatus.

The occurrence of images within the text is currently handled as well. The `<figure>` tag is used, and a free description of the picture can be inserted in a nested `<figDesc>` tag; Leiden+ simply indicates it with the free description preceded by a hash mark. In medical papyri this proves quite useful when dealing with the cases of illustrated herbals, where the extant images can be easily encoded with `#plant` (= `<figure><figDesc>plant</figDesc></figure>`).[134]

4.8 Intertextuality & hypotextuality

Intertextuality is defined as the relation between parallel text, in the form e.g. of quotation or allusion;[135] hypotextuality (with its opposite, hypertextuality) as the relation between a text and a preceding one that is transformed, modified, elaborated or extended. Due to the high degree of both theoretical and practical re-elaboration of medicine – "reperformance", in a sense, to borrow a term created to describe the interplay between text transmission and representation in classical drama[136] –, medical papyri show a complex degree of both inter- and hypotextuality. Not only are the 'classical' medical treatises and handbooks copied following the original text, but they are also quoted, or referred, or re-elaborated in other writings[137] (anonymous treatises as well as manuals, catechisms or collections of prescriptions, and of course commentaries) and excerpted by the late compendiasts (Oribasius, Aetius, Paul of Aegina), who took and interwove excerpts from the earlier authors in order to create composite texts, with the purpose of assembling the best from previous writings.[138]

---

133 On this typology of signs cf. TURNER 1987, 10–12; CRIBIORE 1996, 83–8; COLOMO 2017; AST 2017.

134 P.Tebt. II 679 + P.Tebt.Tait 39–41 (II AD): http://litpap.info/dclp/63596; P.Johnson + P.Ant. III 214 (IV–V AD): http://litpap.info/dclp/64598 (cf. REGGIANI 2018i).

135 Cf. WORTON – STILL 1990; POLACCO 1998; BERNARDELLI 2000; BERNARDELLI 2010, esp. 9–62.

136 FINGLASS 2015.

137 On the concept of intertextuality applied to ancient quotations cf. BERTI 2012, part. 439–46, about ancient historians.

138 Cf. HANSON 1997,296; ANDORLINI 1997a, 19–20.

The interconnection between all such parallel or derived texts is of the utmost importance for evaluating the history of medical science, the dynamics of ancient textual transmission, and the framework of literacy among medical experts, so that an annotation layer that may link the actual text on the papyrus to any relevant related passage in other sources would be most useful.[139] In the cases of papyri preserving 'literary' works (Hippocrates, Galen, etc.), for example, our fragments quite often do provide more genuine readings than manuscript tradition, since they are chronologically closer to the source;[140] they can therefore support some manuscript versions against others, or even preserve previously unattested variants, facts that deserve a particular attention.

A small selection of significant samples will suffice. In P.Oxy. XIX 2221r + P.Köln V 206r (http://litpap.info/dclp/61916), the abovementioned I-century AD commentary to Nicander's *Theriaka*, the extant quoted passages generally agree with the more recent manuscripts of Nicander's tradition (= ω) against the ancient codex Parisinus Π, and show also new genuine variants.[141] The comments, in turn, do not show many points in common with the known scholiastic tradition, and may be traced back to the most ancient comment to Nicander's *Theriaka*, that by Demetrius Chlorus.[142] In the Aberdeen Hippocratic papyrus (GMP I 1), already cited with regards to the adaptations to the Greek language spoken in Egypt, we do find variants already attested in the manuscript tradition (ll. 4–5) but also passages completely divergent from the codices (ll. 11–12, where the length of the gap and the shape of the following traces exclude the unanimous manuscript tradition, which is of course printed in all the editions, in favour of a previously unattested variant).[143]

Alignment among parallel versions of the same text, by linking external resources providing canonical literature,[144] can therefore convey precious information and significant analysis tools, and can be well extended to all the cases (even documentary

---

139 So far, this has been possible only in the line-by-line commentary: cf. BERTONAZZI 2018b and CORAZZA 2018a for discussion and case studies.

140 Cf. ANDORLINI 1997a, 22 n. 15 and 23.

141 E.g. ii,12 πλέει ὄγκος vs πέλει ὄγκος ω & Gow-Scholfield, πέλει ὁλκός Π & O. Schneider (Nic. *Ther.* 387). According to COLONNA 1954, the original text could have been πλέει ὁλκός, subsequently popularized in πέλει ὁλκός and glossed with ὄγκος.

142 Full analysis in COLONNA 1954.

143 Cf. ANDORLINI 2001. For another Hippocratic papyrus preserving an interesting and complex textual history (P.Ant. III 184, VI AD, http://litpap.info/dclp/60192) cf. HANSON 1970; in particular, "the sequences of the Hippocratic texts do not correspond to the one established in medieval tradition but seem to follow autonomous criteria" (CORAZZA 2018b, 174).

144 While digital repositories of literary texts do exist, they usually do not record all manuscript variants of the texts (see above); they could well be connected to the papyrus texts but information would be partial. A possible solution may be to create multitextual digital editions of the literary texts. It is important, of course, to distinguish parallel passages in copies of the same text from quotations embedded in different texts; for the latter, the current platform offers the possibility to deploy the <q>

ones) of texts preserved in more than one item (copies, duplicates).[145] It would also solve the problem of encoding philological variants and manuscript readings in the papyrological digital editions, a challenge faced during the construction of the DCLP database and not yet satisfactorily solved.[146] Alignment of ancient or modern translations of the medical texts (e.g. Latin or Arabic versions) should also be taken into consideration,[147] while a fruitful integration between syntactic annotation/analysis and intertextual referrals may be envisioned for the most intriguing issues of medical literature, as recently argued by Francesca Bertonazzi:

> l'analisi del lessico tecnico dei papiri chirurgici ha portato a individuare paralleli testuali tra testo tramandato su papiro e tradizione manoscritta, talvolta significativamente stringenti come nel caso di P.Strasb. inv. 1187 e diversi passi di Eliodoro *ap.* Oribasio. Alcuni altri papiri (P.Lond.Lit. 166, P.Gen. inv. 111, P.Fuad.Univ. 1, P.Ryl. 3.529), come già notato dagli studiosi, sono caratterizzati da una forte presenza di 'lessico eliodoreo' e da alcune peculiarità proprie del *modus operandi* del chirurgo, come la predilezione di interventi chirurgici che siano il più sicuri

---

tag marking quoted phrases (Leiden+: quotation marks + space), which can be easily used to differentiate the appropriate cases.

145 Traditionally, documentary papyri preserving the 'same' text in multiple copies (for a catalogue of duplicates cf. NIELSEN 2000) are treated in the 'philological' way, i.e. collated and merged in one source archetype: e.g. http://papyri.info/ddbdp/p.tebt;3.1;771dupl (note the suffix 'dupl' added to the URL of the digital text, which advises about the existence of a duplicate of the papyrus). However, a certain degree of uneasiness is felt about such a practice , see e.g. in Jelle Stoop's words: "I disagree with this editorial choice for two reasons. First, in a field like papyrology, every copy of a text deserves full consideration and [...] an archetype that would somehow be considered more authentic than a later copy is an editorial fancy. Copies of the same text, however similar, were written with a purpose in mind, so that edition should be more rather than less interesting. Second, in order to appreciate the fact that we have multiple copies [...], we must ask why different versions of it exist in the first place. The interest of these documents is, therefore, not restricted to the text alone, but extends to the life and afterlife of its copies in relation to one another. In sum, the text of just one fragment does not make for a satisfactory edition of understanding of this [text]. By editing the texts in their own right, we learn about the convention of [...] writing in [Graeco-Roman] Egypt" (STOOP 2014, 185). A new 'phenomenological' consideration of papyrus copies is emerging (cf. YUEN-COLLINGRIDGE – CHOAT 2012, with interesting preliminary comments on textual differences between copies of the same document), but, for now, the digital database is following the 'philological' practice, with a significant loss of information. Giuditta Mirizio (Bologna) is currently working on this topic also from the perspective of digital encoding and XML annotation. On this topic, see also below.

146 The proposed tag (`<app type="variant">`, Leiden+ `|var|`) raised some theoretical and methodological issues, for example whether to choose just one manuscript variant or to encode all possible instances. Moreover, the `<app>` tag (see below) typically envisages a "lemma" (`<lem>`) part, which corresponds to the word(s) in the text, and one or more "reading" part(s) (`<rdg>`), corresponding to the alternative(s) in the apparatus, and it must be clearly thought how this should work in the case of philological variants. Currently, the *Digital Corpus of the Greek Medical Papyri* has adopted the solution to just describe the most relevant manuscript variants in the line-by-line commentary.

147 On translations in the tradition of ancient Greek medical texts see e.g. GAROFALO – FORTUNA – LAMI – ROSELLI 2010.

possibili per il paziente, nonché del *modus scribendi*, come il ricorso frequente alla prima persona – singolare o plurale –, la definizione con esattezza delle posizioni 'topografiche' della parte operata (dentro, fuori, sopra, sotto), e una sostanziale semplicità delle strutture sintattiche usate. Ad oggi, i tentativi di attribuire i papiri citati alla paternità di Eliodoro si sono basati quasi esclusivamente su criteri lessicali nel confronto tra il testo tramandato su papiro e sui capitoli di Oribasio che portano la titolatura 'da Eliodoro'. Una nuova possibile strada offerta dalle nuove tecnologie della papirologia digitale è quella costituita dall'annotazione sintattica dei testi: un'analisi più accurata non solo del lessico, che come è noto è la parte più 'volatile' della lingua, ma delle strutture morfologiche e sintattiche dei passi del compilatore tardo in sinossi con i testi dei papiri, sia pure nella limitatezza delle pericopi testuali preservate, potrebbe gettare nuova luce anche su questo aspetto tra i più incerti quanto stimolanti della ricerca.[148]

Re-elaboration is probably the most striking feature of technical texts, stemming from oral teaching and then continuously adapting their content according to the developments of knowledge. Medical genres like the questionnaire or the collection of prescriptions illustrate this framework at the best, though we do find plenty of cross-references in treatises too.[149] Catechisms (*erotapokriseis*), for example, are clearly derived from and devoted to some sort of oral teaching, as we pointed out above while discussing of their paratextual devices. Yet there exists a considerable similarity with the literary genre of the "definitions", connected with the research and teaching practice of Hellenic medicine and attested in the Greek Pseudo-Galenian treatise *Horoi* or *Definitiones Medicae* (XIX 346–462 Kühn) and in the Latin Pseudo-Soranian *Quaestiones medicinales*.[150] In fact, David Leith has recently distinguished two types of question-and-answer medical texts: the proper catechisms, being introductory manuals for the student of medicine, and wider treatises on remedies. The suggestion came from the similarities detected between *erotapokriseis* on papyrus like P.Turner 14 (http://litpap.info/dclp/63560) and PSI inv. 3783 (http://litpap.info/dclp/63244) and the excerpts from the physicians Herodotus and Antyllus preserved in Oribasius' *Collectiones Medicae*.[151] One may also recall the similarities between the abovementioned P.Oslo and P.Oxy. overlapping questionnaires, or between the surgical catechism P.Gen. inv. 111 (http://litpap.info/dclp/63819) and the treatise known as *Cirurgia Heliodori*,[152] or also between P.Aberd. 11 (http://litpap.info/dclp/63332) and

---

148 BERTONAZZI 2018a, 242–3.

149 Cf. e.g. ANDORLINI 2014. "La possibilità di identificare alcuni papiri con trattazioni di un autore tramandato solo indirettamente inserisce tasselli nuovi nella complessa stratificazione della trasmissione indiretta, soprattutto quando sono i papiri i soli testimoni diretti di autori tramandatici per excerpta e citazioni (Apollonius Mys, Heras, Heliodorus, Herodotus Medicus)" (ANDORLINI 1997a, 22).

150 On which cf. KOLLESCH 1963 and FISCHER 1998 respectively. For general considerations about catechisms on papyrus see also BONATI 2018e and BERTONAZZI 2018a, 57–62 , as well as REGGIANI 2016b.

151 LEITH 2007; cf. already ANDORLINI 1997b, 160.

152 Cf. MARGANNE 1986 and now BERTONAZZI 2018a, 237–8.

P.Ross.Georg. I 20 (http://litpap.info/dclp/63569), two ophthalmological catechisms that certainly derive – with variations – from the same source.[153]

Prescriptions are even more complex (and fluid) in transtextual and hypotextual relations. I have extensively dealt with transmission of ancient medical recipes elsewhere, where I outlined the articulated route from oral compositions and draft transcriptions to professional exchange and collection.[154] Medical prescriptions are fragmentary units, which stem from diagnostic-therapeutic practices and oral knowledge that are recorded on wax tablets (*pinakes*), first kept at the sanctuaries of the healing gods, then collected by leading physicians (namely Hippocrates) in order to build systematic medical repertories.[155] At this stage it is hard to trace any actual intertextual relation, but when – seemingly in the early Roman age – prescriptions start circulating among the physicians, the plot gets intricate. Professional doctors exchange single recipes on papyrus scraps with each other and collect those fragments of medical knowledge into lists and catalogues on parchment booklets, deploying a set of paratextual devices to preserve the unity of each prescriptive text. Galen is the best witness to this 'research' activity,[156] as well as of the 'philological' attention to the pharmacological books held by the libraries, which he himself consulted and collated to get the most exact versions of the texts and to compile his famous treatises on the composition of remedies.[157]

This workflow is by no means exhausted with Galen: among the numerous possible examples, P.Berl.Möller 13 (http://litpap.info/dclp/64268) is a stunning instance. This papyrus, a comparatively large portion of a roll from Hermoupolis Magna, dated between the late III and the early IV century AD, is written on the *recto*

---

153 Cf. Bertonazzi 2018a, 236–7, with earlier bibliography.

154 Reggiani 2018g and 2018j.

155 Cf. Totelin 2009a, part. chapters 1–3.

156 *Comp.med.loc.* I 1 = XII 423,13–15 K. (a recipe is found in a dead physician's parchment notebook and then forwarded to Galen); *Antid.* I 5 = XIV 31,10–15 K. (exchange of recipes); *Indol.* 33–5 (his own personal collection of worldwide prescriptions, destroyed by the AD 191 fire). See also P.Mert. I 12,13–24, attesting to the very same activity of exchange between two colleague physicians in Egypt.

157 The ancient practice of collating several copies (*antigrapha*) of medical texts is attested above all by Galen, who noted several degrees of manuscript divergences, ranging from small linguistic variations to major discrepancies in the content, e.g. in the ingredients and quantities (cf. Andorlini 2000, 38–9; Andorlini 2003, 14–15; Totelin 2009b; Bonati 2016b, 64–5), but we know of other cases in which the ancient readers produced 'personal' copies that became, by means of reformulations and abbreviations, new recensions of the same text (cf. Andorlini 2000, 37–8). In some cases it is possible to speak of erroneous or inaccurate deviations from the original (it is the case with Galen's treatises, for which the ancient author himself stigmatised the circulation of incorrect versions of his own books: *De libris propriis* II 91–3 Müller = XIX 8–11 K.; cf. Hanson 1985, 43–5) but in other cases it is difficult to go back to a genuine text (Hanson 1985, 34–5 makes the example of Hippocratic letters). In general, on Galen's 'philological' work cf. Hanson 1998; Andorlini 2003, 15–16; Dorandi 2014; Bonati 2016b, 63–5.

along the fibres, therefore purposely produced as a collection of medical prescriptions, of which only two columns survive. The first one contains a single prescription "to prevent hair loss on the head", identified by MARGANNE 1980 as a prescription ascribed by Galen to Heras of Cappadocia, a pharmacologist active between 20 BC and AD 20. The text on the papyrus parallels Gal. *Comp.med. loc.* XII 430,8–15 K. verbatim,[158] while other variant versions of the same remedy are recorded by Galen himself (*ibid*. XII 435–6 K.) as antecedents of Heras' one.[159] Subsequently, CORAZZA 2016 discovered that also some remnants of the second column can be identified with other recipes by Heras, this time against headache, mentioned by Galen as well, with some wording variants.[160] Two of them patently parallel Galen, but the papyrus is by no means a copy of *On the composition of medicaments by places*: the recipes do not follow the canonical order in which they are cited in Galen's treatise, clearly attesting a work of selection, extraction, and thematic re-arrangement, in which each recipe is treated as a unit to be managed on its own; moreover, the other two identified prescriptions look like variants of Heras' texts as reported by Galen, thus attesting a 'fluid' stage of transmission, in which recipes are modified and adapted according to the users (Galen himself, as we saw, attests some earlier versions of Heras' recipe against hair loss). It is apparent that this interconnection of "living texts",[161] copied and re-copied from original pieces or different collections, generates cross-references and inter-quotations that may well fall into the cases described in these paragraphs.[162]

4.9 Metatextuality

Metatextuality is the explicit or implicit critical commentary of one text on another text. For the same reasons described above apropos of paratext and intertextuality, namely the fluidity of medical technical texts, always subject to renovation and up-

---

158 In fact there are some interesting variants, which as usual show how papyri can contribute to the history of the texts: in particular, at line 10 (καλοῦσι pap. : καλοῦσι καί Gal.) the papyrus offers a superior reading, since the conjunction is syntactically unfit; further discussion in CORAZZA 2016 *ad locc*. On the value of the variants attested in the papyri see above.

159 Cf. MARGANNE 1980, 182–3.

160 In particular, the first prescription of the second column (ll. 1–3) parallels Gal. *Comp.med.loc.* XII 593,14 K. verbatim, while the following two (ll. 4–8 and 9–15) show partial overlaps with *ibid*. XII 594,1–4 (= Aet. VI 50,75–9) and XII 594,7 ff. K. All these recipes are ascribed to Heras. The remaining traces of fifteen lines, articulated in four more recipes, could not be identified with any known text.

161 BONATI 2016b, 66.

162 The Leiden+ tag for parallel passages is meant to mark omitted text that is supplied on the ground of parallels, so it does fall into a different typology (see below, editorial interventions).

date after practical and individual experience, the practice of annotating is widespread in the medical papyri.[163] The annotations can acquire either the organic format of the commentaries, autonomous exegetic treatises, the most illustrious examples of which are Galen's commentaries on Hippocratic texts[164] (see e.g. Gal. *In Hp. Epid.* II 4 = CMG V 10,2,1, p. 78,7–11 Wenkeback, where the author himself explains some reasons for compiling commentaries), or the scattered aspect of the *scholia* or marginal annotations: see the examples of P.Ant. I 28 (http://litpap.info/dclp/60189), fragment of a III-century parchment codex from Antinoupolis with the text of Hippocrates' *Aphorisms* and *marginalia*,[165] as well as of P.Ant. III 186 (http://www.litpap.info/dclp/59961), a very fragmentary large-format papyrus codex from the same place, dated to the VI cent. AD, which contains sections of Galen's *De compositione medicamentorum per genera* along with some scanty marginal annotations.[166]

In both cases, textual relations are complex.[167] Commentaries refer to other texts but without exact parallels, except for literal quotations (see above); marginal notes refer to the main text without being part of it, so that the current treatment in digital editions may be slightly misleading, since it allows for marking the marginality of the passage (added to or written into the margins), but not the type of relation with the main body of the text.[168] The XML syntax is clear: *marginalia* are encoded as plain text lines, with the indication of the margin attached to the line number.[169] Let us consider, instead, a more complex case, represented – again from Antinoupolis – by P.Ant. III 126 (http://litpap.info/dclp/65233):

> P.Ant. 3.126 (VI–VII secolo d.C.) è parte di un compendio sul trattamento farmacologico e chirurgico della tonsillite e rappresenta un esempio di 'enciclopedia medica' redatta in epoca bizantina; il ritrovamento di testi come questo conferma l'idea che nella pratica medica antica la trasmissione del sapere avvenisse tramite la combinazione di fonti tradizionali, tramandate per tradizione scritta, e di materiale desunto dalla pratica medica quotidiana e registrato proprio dagli specialisti che operavano sul campo.
> Il testo principale, ovvero quello scritto in carattere più grande nella parte più estesa di papiro, è arricchito da annotazioni nel margine inferiore che riguardano alcune terapie farmacologiche da impiegare nel caso dell'insorgere delle patologie descritte nel testo, e tale modalità di uso e

---

163 On the practice of annotating medical treatises with *scholia* and comments cf. ANDORLINI 2003; in general, on *scholia* and commentaries in the papyri cf. MESSERI SAVORELLI – PINTAUDI 2002..

164 Cf. MANETTI – ROSELLI 1994.

165 cf. ANDORLINI 2000, 41–2; ANDORLINI 2003, 19–24

166 Cf. CORAZZA 2018a.

167 The two cases are tightly related, and can even merge together in the so-called "commented editions" discussed by VANNINI 2015.

168 See the observations by CORAZZA 2018a. On the interactions between text and glosses, very interesting is the analysis by MANIACI 2002, though referred to later types of texts.

169 E.g. `<lb n="1,minf"/>` for lines in the bottom margin (the other margins are indicated with `msup`, `ms`, `md`). In Leiden+ this information is added to the line number accordingly.

riuso del testo testimonia l'*iter* con cui il sapere tradizionale era compendiato, arricchito e integrato nei libri tecnici dai possessori dei testi. Le caratteristiche di layout, la consistenza dei margini (quello inferiore, quasi totalmente conservato, misura 5 cm) e la scrittura regolare, oltre all'indicazione in alcuni punti degli spiriti, lasciano pensare che il frammento fosse parte di un codice di notevoli dimensioni e, dunque, di un certo pregio; il tipo di annotazioni riportate nel margine, anche in mancanza di notizie più specifiche circa l'uso di questo codice, fanno pensare che il redattore potesse essere un medico piuttosto competente o un soggetto forse ancora in formazione ma abituato alla pratica medica.[170]

The relation between the *marginalia* and the text is tight, though the current markup can be arranged just as follows:

P.Ant. 3.126, Fr. A

13,minf. [α]ζυμου ε .1 [.4]λων του ξηριου [.15] [ε]πιχεε θερμον
14,minf. [υ]δωρ και λει[οτριβε] καλως μετα .1 [.15] [ε]λαιου καταπλασσε
15,minf. [.4]αρτου .1 [.7].2 ουσπε .3 [.?] [πλ]ασσων το κατα
16,minf.– [πλασμα] [.?]
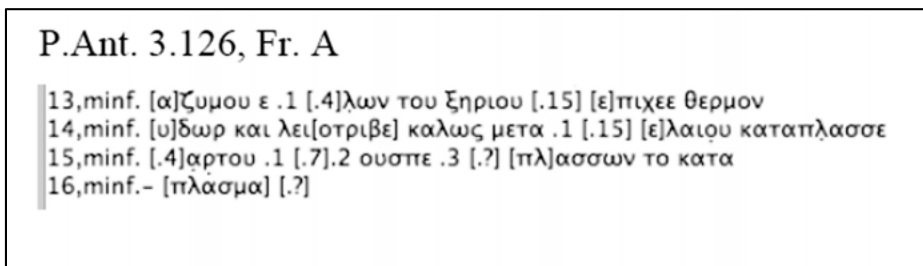
Fig. 7: Sample markup of *marginalia* according to the current standards (from Bertonazzi 2018a, 73).

It is clear that we are not dealing with simple additions to the text, which are easily encoded with the `<add>` tag, further specified – with the `place` attribute – according to the position of the insertion (above, below, left, right, interlinear).[171] Scribal additions can be effectively utilized under particular circumstances, as in the case of P.Oxy. IX 1184v (http://litpap.info/dclp/60175), a I-century AD fragment exhibiting part of a collection of Hippocratic epistles likely arranged by theme (the extant texts deal with Hippocrates' invitation to Persia by the Great King, which he self-confidently refuses).[172] The papyrus contains different versions of the Pseudo-Hippocratic letters 3, 4, 4a, 5, and 6a (ed. Smith), separated by initial *ektheseis*, and *paragraphoi* between each other.[173] *Ep.* 3 is shortened at the end, and its 'canonical' conclusion has

---

170 Bertonazzi 2018a, 53–4; cf. also *ibid.*, 73 for its digitization; Corazza 2018b, 46–57; on the annotations, McNamee 2007, 463 ff.

171 For the cases of scribal additions, Leiden+ recovers some traditional Leiden conventions, so that supralinear insertions are encoded between two slashes (\x/) and infralinear insertions with reversed slashes (//x\\). The other types of additions are rendered as ||left:x||, ||right:x||, ||interlin:x||.

172 Cf. Brodersen 1994, 103–7.

173 The papyrus presents also interesting cases of intertextuality (see above): of *Ep.* 5, it transmits the shorter form with certain variations, while *Ep.* 6a, a letter to Gorgias previously unattested, has striking coincidences of phraseology with 'canonical' *Ep.* 6, addressed to Demetrius.

been appended as a supralinear insertion flowing into the right-hand margin – this is easily encodable with a combination of the two relevant `<add>` tags. But then *Ep.* 4 was transcribed twice, in an abridged version in the main text, flanked by a shorter form without the introductory salutation (*Ep.* 4a), added into the right-hand margin and separated from the main body of letter 4 with an irregular vertical line. Further below, between letters 4 and 5 (ll. 17–19), three lines of comment appear, unattested elsewhere.

Marginal or interlinear additions merge with comments in a complex metatextual net that sometimes overflows into the text itself[174] and show a remarkable 'philological' care for the text by the ancient scribes. The case of Hippocrates' fourth letter, described just above, is rather meaningful: the marginal text is not a comment (like the following interlinear insertion) nor an addition (like the preceding supplementary insertion), it is an alternative parallel version, a proper variant of the text presented in the main body of the papyrus. In this case, the vertical, irregular line traced by the ancient scribe to divide the two alternatives acts as a proper indication of a textual variant.[175] We do find even more puzzling instances. P.Tebt. II 272v (http://lit-pap.info/dclp/60048, late II cent. AD) is a fragment of Herodotus Medicus' *De Remediis*, describing the symptomatology of thirst and its treatment; the text corresponds in part to an excerpt of Herodotus Medicus preserved with Oribasius' treatment of thirst in case of fever (*Coll.Med.* V 30, 6–7 Raeder = CMG VI 1,1). At a certain point, where the text reads αἰτίαι τῆς προσφορᾶς (l. 5), introducing the different reasons for giving the sick something to drink, the scribe added two groups of three letters between dots above the line:[176] *τῶν* above τῆς, and *ρῶν* above ρᾶς. This is not an addition *supra* or *infra lineam*, since it is clearly an alternative to the syntagm below (plural instead of singular); and since nothing appears deleted, it is not clear if the ancient writer wanted to correct the text or just juxtapose two different versions of the same passage.[177] We cannot be sure of what is going on here because this variant is unattested in the manuscript tradition, i.e. in Oribasius' passages quoting Herodotus Medicus, which all feature the singular form. We would have a scribe correcting the form unanimously preserved by the manuscript tradition and replacing it with an

---

[174] Sometimes a marginal annotation can be swallowed up in the main text, generating a textual issue that can be likely explained only by means of metatextual correlations: this is the case with P.Gen. inv. 111 (http://litpap.info/dclp/63819), where the reading ῥάμματος ἢ μί|[τ]ου (ll. 15–16), presenting two technical terms that are almost synonyms, may stem from a gloss (cf. BERTONAZZI 2018a, 241–2).

[175] It is worth noting that similar graphical devices are used by the author of the *Anonymus Londinensis* to frame alternative versions of the same passage (cf. CRIBIORE 2018 and see below).

[176] I thank very much Todd M. Hickey and Derin McLeod for the help in getting a high-resolution picture of the fragment. I mention this case in REGGIANI 2018a, 2018b, 2018c, with discussion of the tentative code used to digitize it.

[177] Writing a word between dots could be a way to highlight later corrections, like e.g. the *koppa* in P.Eirene III 25, 3 (III AD; see comment *ad loc.*).

unattested variant. The P.Tebt. editors speak of "correction or alternative reading", M.-H. Marganne of "hésitation";[178] if we should define it, we ought to call it a 'scribal variant', just as in the Hippocratic case presented above, as well as in P.Oxy. LVI 3851 (http://litpap.info/dclp/61917, II–III AD), a fragment of Nicander's *Theriaka* (333–4), which at l. 12 reads πρεσβίστατ[ον] (attested in most of the manuscript tradition) with a υ added *supra lineam* between dots, being πρεσβύστατον an alternative version attested in some of the manuscripts (= Kv).
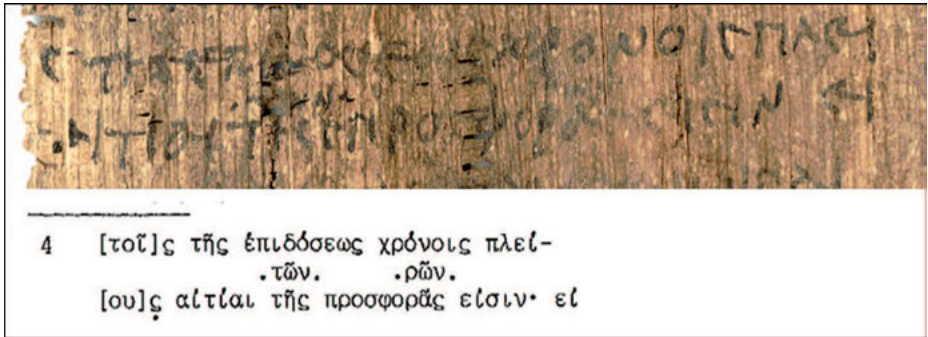
Fig. 8: P.Tebt. 272,4–5 (courtesy of the Center for the Tebtunis Papyri, University of California, Berkeley).

It is not easy to deal with these cases digitally,[179] at least with the currently available tools, which deploy instead a full set of tags aimed at encoding plain scribal corrections, i.e. additions (see above), deletions (`<del>` tag with `rend` attribute describing the type of deletion: `"erasure"`, `"slashes"`, `"cross-strokes"`),[180] and replacements (`<subst>` tag containing a nested `<add place="inline">` tag defining the corrected text and an equally nested `<del rend="corrected">` defining the replaced text).[181]

---

178 Marganne 1981b, 76.

179 Tomasi – Zaja 2002 discuss some interesting solutions for the encoding of marginal writings, though dealing with quite later types of texts.

180 Leiden+ employs the double square brackets as in conventional printed editions; only square brackets for plain erasures, ⟦/ ⟧ for slashes and ⟦X ⟧ for crosses.

181 Leiden+ employs a |subst| tag working the same wasy as |reg| and |corr|. An interestingly complex case (P.Strasb. inv. 1187, A, i,11 = http://litpap.info/dclp/59968) is presented by Bertonazzi 2018a, 64 and 69–70: an ancient scribal correction, involving the insertion of a letter *supra lineam*, was read differently by two editors, so that they proposed two different interpretations, one of which involved a regularization. The newer reading is σμειλιοτῶν corrected in σμειλι\ω/τῶν i.e. σμιλιωτῶν; the previous reading was νω δεῖ λιοτων corrected in νω δεῖ λι\ω/των. In Leiden+ this is encoded as <:<:<:σμιλιωτῶν|reg|σμ̣ε̣ιλι\ω/τῶν:>|subst|σμ̣ε̣ιλιοτῶν:>|ed|νω̣ δεῖ <:λι\ω/των|subst|λιοτων:>=ed.pr.:>, and the XML is cross-nested accordingly, so that the current display on the platform appears quite messed up.

The philological care testified by the cases of 'scribal variants' mentioned above is even more patent when the text is an autograph[182] and is equipped with authorial revisions, for which an important contribution can come from the XML annotation of genetic criticism phenomena recently developed by Elena Pierazzo.[183] Raffaella Cribiore has recently showed how genetic criticism – aimed at reconstructing the process of authorial constitution of a text – can be successfully applied to papyrological texts.[184]

4.10 Editorial interventions (modern)

Modern alternative readings and editorial supplements do influence linguistic annotation, in that they add data, which are not *stricto sensu* 'original' to the text. Alternatives produce multiple possible readings, one of which is usually the most probable but without full certainty, and the other possibilities may well fit the context. Supplements, though most likely and in some cases pretty unavoidable, are nonetheless a modern contribution to the ancient fragmentary text and deserve a particular attention. They can even be incorrect, and thus fall into the third category of editorial corrections, which encompass all modern corrections made to the readings of previous modern editors. Alternatives and editorial corrections are currently encoded as apparatus elements (`<app>`) defined by the `type` attribute and composed of a "lemma" (`<lem>`), i.e. the word in the text, and a "reading" (`<rdg>`), i.e. the alternative in the apparatus. In Leiden+ they are marked with the `|alt|` and the `|ed|` tag respectively. Such a markup strategy works fine from the 'philological' viewpoint, since it provides a main reading – supposedly the most correct – and a set of critical alternatives, either proposed by the same editor or sedimented through years of scholarship, with the possibility to indicate the authorial responsibility for each reading (`resp` attribute in the `<lem>` tag).[185] Nevertheless, the impossibility to search for combination of words including the terms in the apparatus makes this choice rather uneasy for the purposes of digital databanks, while different layers of text, each one featuring a single textual

---

182 On medical autograph papyri cf. Andorlini 1997a, 22 with earlier bibliography; Marganne 2004, 90–1.

183 Cf. http://www.tei-c.org/Activities/Council/Working/tcw19.html; Pierazzo 2008.

184 Cribiore 2018: see in particular the case of the medical *Anonymus Londinensis* and the related discussions of double versions. From the computational viewpoint, cf. Macé – Baret – Bozzi – Cignoni 2006 (in particular, Passarotti 2006). Genetic criticism can be applied to some documentary categories which show a certain complexity of textual composition. One may recall, just for instance, the legal documents of Ammon's archive, produced in multiple versions (P.Ammon II; cf. Cribiore 2018); Raffaele Luiselli's considerations about authorial revisions in Roman letters and petitions (Luiselli 2010); the mostly neglected cases of duplicates recently 'rediscovered' by Malcolm Choat and Rachel Yuen-Collingridge (Yuen-Collingridge – Choat 2012); the composing process of administrative reports studied in the Project *Synopsis* at the Heidelberg University especially by Uri Yiftach (cf. Reggiani 2016c); the very recent discussion on drafts and copies by Andrea Jördens (Jördens 2017).

185 See, however, Damon 2016 for criticism of this way of handling apparatus readings by TEI.

alternative, may enhance the digital representation of the papyrus, especially when considering that editorial interventions occur more frequently in the medical papyri than in the documentary ones, for the peculiar attention to the editorial history that characterizes the items of the *Digital Corpus of the Greek Medical Papyri*.[186]

On the other hand, supplements are tagged as such with the `<supplied>` element, and a `reason` attribute that defines the type of integration: `"lost"` if the original text is lost (Leiden+ square brackets), `"omitted"` if the original text was left over by the ancient scribe (Leiden+ angle brackets), `"parallel"` if the text is inserted on the ground of a parallel text (Leiden+: pipes + underscores |_ _|). The opposite case (removal of ancient surplus text) is marked with a `<surplus>` tag (Leiden+: curly brackets). In this case, integration with the text layer is granted by the fact that the `<supplied>` tag indicates a text portion. This is even clearer if we compare it with the tag used to mark unsupplied lacunas, that is a `<gap>` tag with a `reason` attribute set to `"lost"` (see above).[187]

The case of the gaps is indicative of the semantic difference between digital and paper edition. In a printed critical system, both supplied and unsupplied lacunas are marked with square brackets because the focus lies in the descriptive layer of the papyrological fact: a certain missing part of the text, which may be recoverable or not. In a digital critical context, we need to define whether a lacuna bears a textual meaning (i.e., a supplied text) or not (i.e., a gap in the text). Leiden+, following the printed conventions, adopts square brackets for both, in order to help the users; but the system automatically chooses the appropriate XML code according to the content of the brackets. Therefore, when the papyrus displays a partially supplied gap, which is enclosed by the very same pair of brackets in the printed edition, in the digital edition the two different parts (supplied and unsupplied) must be kept separated since they mean two different facts. Leiden+ brackets are different than Leiden printed ones also in that the former must be always opened and closed at each gap, while in a printed edition they can be left open (or unclosed) if their exact extent is unknown.

Somehow ambiguous, in conclusion, is the treatment of modern corrections in the case of misspelled words. Though the typical treatment involves the |reg| and

---

186 Cf. CORAZZA 2018a; BERTONAZZI 2018a, 64–7 (with case studies) and 2018b.

187 The theoretical assumption that the fragmentary status of the papyri may be thought as a (paradoxical) sort of 'non-voluntary quotation', selected by the chance and by the material circumstances rather than by the will of an author, would allow to envision a transtextual link between a 'virtual' hypertext (the original document, lost, more or less recoverable in a philological way) and the concrete hypotext (the actual fragment; cf. REGGIANI 2016a; see also ROMANELLO – BERTI – BOSCHETTI – BABEU – CRANE 2009, 160 and 162: "[...] fragments do not actually exist outside of scholars' interpretations. [...] Fragments are always scholarly reconstructions and interpretations of the content and structure of lost works"). This may allow for creating several multiple layers for editorial alternatives and supplements too, thus avoiding complicated nested tags as in the case of multiple alternatives in a series of different supplements or modern editorial readings (see the samples provided in the Conclusions below).

|corr| tags according to the type of intervention (see above), the use of traditional Leiden conventions (angle brackets for supplements of omitted letters, curly brackets for removal of surplus letters) is admitted in case of outright diplographies ("where the letter(s) is genuinely superfluous", so say the guidelines)[188] or trivial omissions (for which it is preferred to the |corr| tag).

4.11 Image

When available, the addition of a digital picture is fundamental for a complete evaluation of the papyrus. The advanced possibilities of virtual objects, of which I discussed elsewhere,[189] could be further enhanced by aligning text and image, a procedure that has been successfully attempted by the *Anagnosis* project at Würzburg.[190]

5 Concluding remarks

The so-called *Michigan Medical Codex* (P.Mich. inv. 21 = P.Mich. XVII 758, http://www.litpap.info/dclp/59332)[191] resumes at the best most of the preceding arguments. It is a IV-century small-format papyrus codex, of which thirteen leaves survive to an amount of twenty-six pages, in which numerous recipes are collected – seemingly – according to type of medication (pills and lozenges, then wet and dry plasters, at least in the extant pieces). Commissioned by a practicing physician,[192] the document shows various degrees of textual interventions. In the original writing, recipes start with an indented heading, declaring the type of remedy, and are separated from each other with lines and small blank spaces; they typically contain the list of ingredients, followed by directions for composition and use. Many prescriptions are ascribed to famous doctors, showing

> correspondences with recipes for plasters in the collections of Galen, Oribasius, Aëtius, or Paul of Aegina that have come down in the manuscript traditions, highlighting the striking degree of continuity among ingredients and their relative proportions from hand-written copy to hand-written

---

188 For two cases in the medical papyri, see BERTONAZZI 2018a, 67–8 ({τῶν σιναρῶν} in P.Strasb. inv. 1187, A, i,14 and {σ}σχηματίσαντες in P.Lond.Lit. 166, iv,6).

189 REGGIANI 2017, 137 ff.

190 See above and the *Anagnosis* section of R. Ast's and H. Essler's chapter in this volume.

191 YOUTIE 1996. For the following observations, I refer to HANSON 1996, HANSON 1997, 302–4, and ANDORLINI 2003, 26–8.

192 Cf. YOUTIE 1996, 1–3; ANDORLINI 2003, 26–7.

copy over many centuries: the presence of plasters from a variety of different physicians suggests that the basic text of the codex was combining and taking its shape over considerable time.[193]

Then, the interventions by the owner of the codex:

> First he collated the text of his newly-made copy against an exemplar, making corrections in addition to the items already corrected by the scribe, and then he went on to more than double the contents of the codex by filling the margins with additional recipes for pills to medicate bodily ills and plasters to medicate wounds and lesions of every kind. Because empty space was limited, he emphasized separation between recipes through lines and marginal markers.[194]

Intertextuality, hypotextuality and similar connections merge together, creating a very complex and unique clockwork: "although individual recipes in a collection on papyrus often resemble items in the known authors, each extensive collection on papyrus has thusfar proved to be a unique assemblage".[195] The paratextual function of critical and lectional marks stresses the "composite" structure of the text,[196] while authorial corrections and phonetic variants are not absent from the textual level.

Let us compare a part of the printed edition with the corresponding digital edition currently featured in the DCLP,[197] followed by a tentative proposal of (partial) ontology network to describe the multiple textuality of the sample.

Fig. 9: P.Mich. XVII 758 H r/v: main text with recipes taken from other authors; marginal annotations and additions with a reference system of coronides and other graphical marks (Youtie 1996, Pl. 8).

---

193 Hanson 2010, 197–8, and 1997, 303.

194 Hanson 2010, 197–8, and 1997, 303.

195 Hanson 2010, 199; cf. also the observations by Bonati 2016b, 60–9.

196 Cf. Andorlini 2003, 26–7.

197 The DCLP digital edition of the *Michigan Medical Codex* has been encoded by students of the Papyrology class (F. Bertonazzi, F. Corazza, L. Rizzardi, M.E. Galaverna, C. Bottioni, M. Catania, F. Giraldi, P. Lillo, G. Saccani, E. Mazzetti, L. Mazzolari, A. Brunazzi, E. Angolani, N. Pajares Collado, C.M. Ferrari) under the supervision of L. Iori, M. Centenari, I. Bonati, and N. Reggiani.

```
            - - - - - - - - - - - - -
            .[.]..[
            ἐνώϲα[ϲ] ϑυεί[α
          ⎰αὐτὴν [..].δοτε[
    4     ⎱ἡ ἔμπλαϲτροϲ ῥοδ[ίνη  ±3 ]
            καθάραϲ τὰ πέταλ[α ὑϲϲώ-]
            που κλάδου τοῦ δ[

(Hd.2)    [ἔ]μπλαϲτοϲ ἀφοδευ[ο]μένῳ ϲαϲϲ[ομένῳ  ± 15]
    8     ⎰μοι τῷ πόνῳ .....[
          ⎱δέδοκ(α) κηροῦ (οὐγ.) α .[
     ⍴  ┿ μίϲυος εϲβ( ) (δρ.) α α.[
          ⎱χαλβάν(ηϲ) (οὐγ.) β μυ[
   12     χαλβάνην καὶ [
            [[...αϲαϲαρ]] ϑυ[
            διηθήϲαϲ χρῶ [

9. δέδωκα, Pap. δεδοϰ    10. Pap. εϲβ    11. Pap. χαλβαν
```
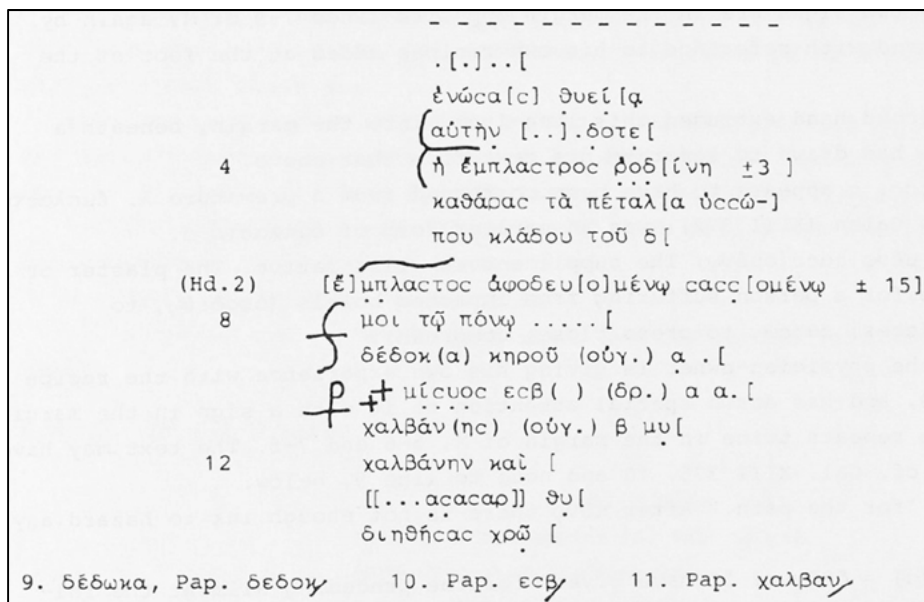
Fig. 10: Printed edition of P.Mich. XVII 758 H r (Youtie 1996, 59).

```
P.Mich. Inv. 21 H verso                          9 x 8 cm.
            - - - - - - - - - - - - - -         Plate 8b
                    ]καὶ π[ρ]ὸϲ τοὺϲ
            [κόλπουϲ προϲφά]του[ϲ] καὶ τῶ(ν)
            [ἑλκῶν ϲυνουλο]υμένων καὶ          ξ (ηρόν)
    4       [πρὸϲ τὰ δυϲ]επούλωτα ἕλκη          πληρ(οῖ)
            [καὶ πρὸϲ δ]ὲ τὰϲ χιμέτλαϲ κὲ

(Hd.2)    [ἔμπλαϲτροϲ ἡ διὰ πο]μφόλυγοϲ πομφόλυγοϲ (οὐγ.) β
          [λιθαργύρου (οὐγ.) γ ϲτέατοϲ ὀρ]νιθεί[ου] (οὐγ.) β κηροῦ (οὐγ.)δ
    8     [χυλοῦ λινοϲπέρμου (οὐγ.). χαμαιμήλου] (οὐγ.) ε τὰ τηκτὰ
          [κατὰ τῶν λείων τῆϲ λιθαργύρου κ]αὶ τοῦ πομφόλυγοϲ

2. Pap. τῶ    3, marg. Pap. ξ|    4, marg. Pap. πληρ    5. χιμέτλαϲ:
τ corr. fr. ϑ, καί
```

Fig. 11: Printed edition of P.Mich. XVII 758 H v (Youtie 1996, 61).

folio H

r

-- -- -- -- -- -- -- -- --
1   ˙[˙]˙˙[-ca.?- ˙˙˙]
     ἐνώσα[ς](*) θυεί[α˙](*) [-ca.?-]
     αὐτὴν [˙˙]˙δοτε̣ [-ca.?-]
     ————
     ἡ ἔμπλαστρος ῥοδ[ίνη - ca.3 -]
5   καθάρας(*) τὰ πέταλ[α ὑσσώ-]
     που κλάδου τοῦ δ[-ca.?-]
     ————
     (hand 2) [ἔ]μπλαστος(*) ἀφοδευ[ο]μένῳ σασσ̣[ομένῳ - ca.14 - ˙]
     (hand 2) (added at left: ((coronis))) μοι(*) τῷ πόνῳ ˙˙˙˙[-ca.?-]
     (hand 2) δέδοκ(α)(*) κηροῦ (οὔγγ) α̣ [-ca.?-]
10(hand 2) (added at left: χρ(ηστός)) ((cross)) (*) μίσυος ἔσβ(εσα(?)) (δραχμ) α α̣ [-ca.?-]
     (hand 2) χαλβάν(ης) (οὔγγ) β μυ[-ca.?-](*)
     (hand 2) χαλβάνην καὶ [-ca.?-]
     (hand 2) [˙˙˙ασασαρ] θυ[-ca.?-](*)
     (hand 2) διηθήσας χρῶ [-ca.?-]
     -- -- -- -- -- -- -- -- --

v

     -- -- -- -- -- -- -- --
1   [-ca.?-] καὶ π[ρ]ὸς τοὺς
     [κόλπους προσφά]του[ς] καὶ <ἐπὶ(?)> τῶ(ν)
     [ἑλκῶν συνουλο]υμένων(*) καὶ
     [πρὸς τὰ δυσ]επούλωτα ἕλκη
5   [καὶ πρὸς δ]ὲ τὰς χιμέτλας(*) κὲ(*)
3,md(hand 2) ξ(ηρόν)(*)
4,md(hand 2) πληρ(οῖ)
     ————
     (hand 2) [ἔμπλαστρος ἡ διὰ πο]μφόλυγος πομφόλυγος (οὐγκ) β
     (hand 2) [λιθαργύρου (οὐγκ) γ στέατος ὀρ]νιθεί[ου] (οὐγκ) β κηροῦ (οὔγγ) δ
     (hand 2) [χυλοῦ λινοσπέρμου (οὐγκ) 1 χαμαιμήλου] (οὐγκ) ε τὰ τηκτὰ
     (hand 2) [κατὰ τῶν λείων τῆς λιθαργύρου κ]αὶ τοῦ πομφόλυγος

△ H.r.1-2. or [συν]|ἐνώσα[ς]
△ H.r.2. or εὐερ[γῶς]
△ H.r.4-5. or [-ca.?- ἀνα]|καθάρας, or [-ca.?- δια]|καθάρας
△ H.r.7. (hand 2) [ε]μπλαστοσ corr. ex [ε]μ̣[λ̣]αστοσ, l. ἔμπλαστρος
△ H.r.8. or [ἒ]|μοι
△ H.r.9. l. δέδωκ(α)
△ H.r.10. or λ(άμβανε(?)), or λ(άβε(?))
△ H.r.11. or μυ[ρίκης -ca.?-], or μυ[ρσίνης -ca.?-]
△ H.r.13. or θυ[εία -ca.?-]
△ H.v.3. or [ἐπουλο]υμένων
△ H.v.3,md. or ξ(ηρίον)
△ H.v.5. corr. ex χιμέθλας
△ H.v.5. l. καὶ

Fig. 12: DCLP digital edition of P.Mich. XVII 758 H r/v (http://www.litpap.info/dclp/59332).
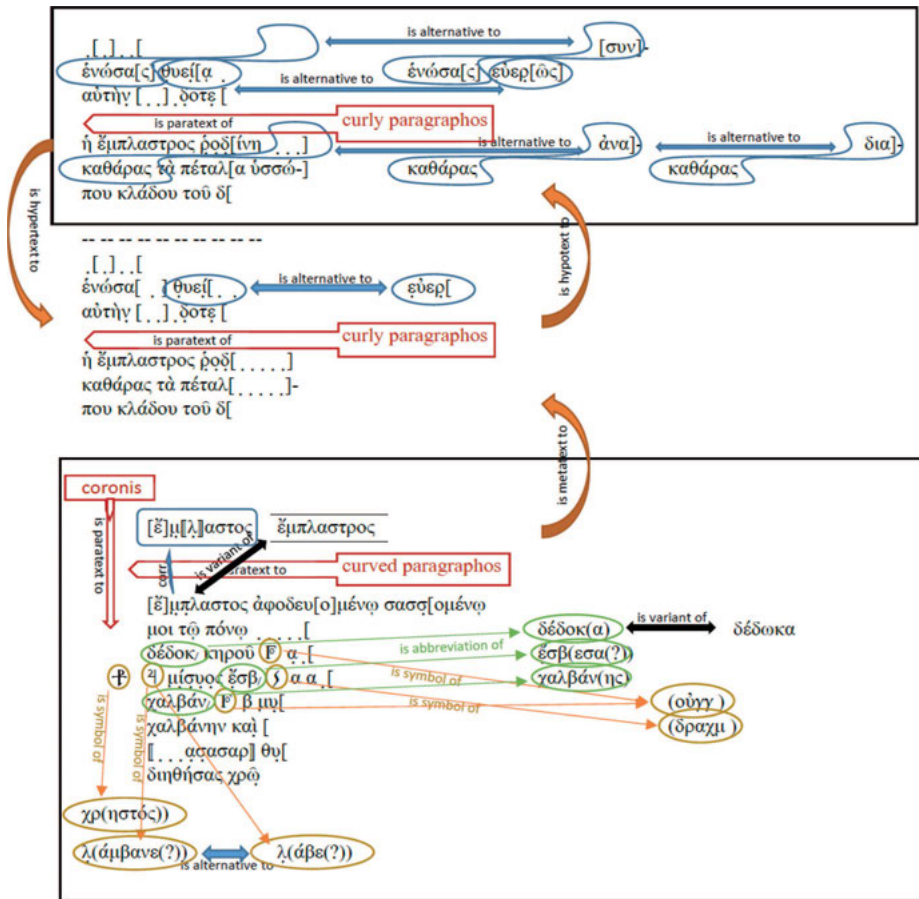
Fig. 13: Tentative ontology model for P.Mich. XVII 758 H r. Some layers are simplified; note that in the metatext layer (below) the hypotext and the hypertext are merged (and some editorial supplements and alternatives are missing) in order to give space to the annotation of abbreviations and symbols, which clearly shows their intensive deployment by the second hand (= the owner of the codex). Noteworthy is [ἔ]μπλαστος (corrected from [ἔ]μ[[λ]], l. 7), which is a substandard spelling variant of the more common ἔμπλαστρος: as already noted by Youtie 1996, *ad loc.*, "ἔμπλαστος, according to Galen (XIII 372 [K.]), was an earlier form of ἔμπλαστρος". Moreover, the entire metatextual paragraph must be noted as written in *ekthesis* in the bottom margin, which is marked line by line in the Leiden+ code. Here, it would suffice to encode it as a marginal metatext and to connect it to an *ekthesis* paratext layer (see the following sample).
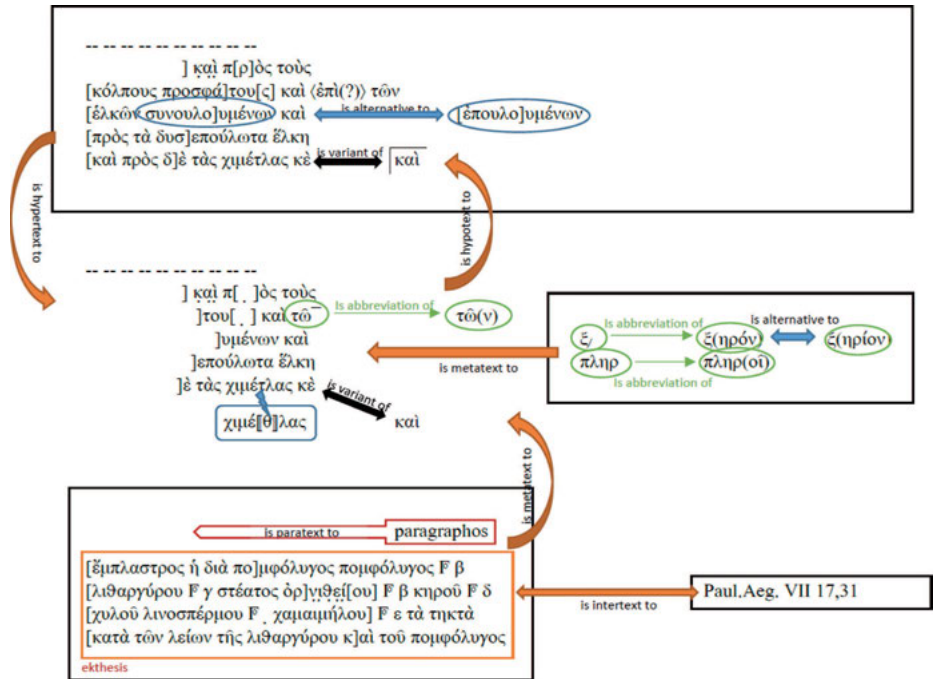
Fig. 14: Tentative ontology model for P.Mich. XVII 758 H v. Some layers are simplified; in the metatext layer (below) the hypotext and the hypertext are merged as in the preceding sample, but here symbols are not handled in order to give space to the *ekthesis* paratext layer and to the intertextual layer, since the recipe added to the bottom margin (ll. 6–9) closely recalls (in the typology and number of the ingredients and in their quantities) a passage of Paul of Aegina (VII 17,31). Note also how the right-hand-margin additions are handled as metatext layer connected to ll. 3–4 of the main text, whereas the Leiden+ markup does not handle the situation properly (marginal lines can be added within the text, or at the end, but in both cases some metatextual information gets lost). Quite interestingly, the scribal phonetic correction χιμέτλας for χιμέ[[θ]]λας (l. 5) attests to a preference for a form used by Paul (e.g. III 79,1) rather than other medical writers (e.g. Orib. *Coll.* IV 615,19; 620; *Syn.* VII 45; Gal. XIII 380,5 K.; 383,17 K.).

Admittedly, printed or printed-like media are physically limited as to dealing with complex degrees of textuality, and adopted the critical edition model as a way of fixing a text for scholarly purposes. On the contrary, ancient textual criticism – recognizable in the commentaries, the annotations, the philological interventions, the paratextual care deployed by the ancient scribes and scholars – was apparently a way to pass down knowledge, i.e. a means of text transmission rather than text reconstruction and fixation. Nowadays, thanks to the digital tools, we do have the occasion to develop digital infrastructures in a hyper-dimensional cyberspace to overcome traditional criticism and its shortcomings, and to conceive a digital critical edition with

deeper and deeper levels of text analysis (markup tagging, linguistic or semantic annotation layers, in-text information).[198]

As BODARD – GARCÉS 2009 argue, a major advantage of digital editions (namely the papyrological ones) is the possibility to get back to the materiality of texts, avoiding the philological necessity of reconstructing an archetype and focusing on text transmission instead. "[A]ttention would be better focused on how to present a text with multiple manuscript witnesses to a reader in a digital environment":[199]

> Digital editions may stimulate our critical engagement with such crucial textual debate. They may push the classic definition of the 'edition' by not only offering a presentational publication layer but also by allowing access to the underlying encoding of the repository or database beneath. Indeed, an editor need not make any authoritative decisions that supersede all alternative readings if all possibilities can be unambiguously reconstructed from the base manuscript data, although most would in practice probably want to privilege their favoured readings in some way. The critical edition, with sources fully incorporated, would potentially provide an interactive resource that assists the user in creating virtual research environments.[200]

> Thus, the authors hop[e] that digital or virtual research environments would support the creation of 'ideal' digital editions where the editor does not have to decide on a 'best text' since all editorial decisions could be linked to their base data (e.g., manuscript images, diplomatic transcriptions).[201]

Similarly, NICHOLS 2009 states that the ideal of the archetype text and textual criticism is an "artefact of analogue scholarship" consequent to the limitations of the printed pages. Conversely,

> [t]he Internet has altered the equation by making possible the study of literary works in their original configurations. We can now understand that manuscripts designed and produced by scribes and artists – often long after the death of the original poet – have a life of their own. It was not that scribes were 'incapable' of copying texts word-for-word, but rather that this was not what their culture demanded of them. [...]. [I]t requires rethinking concepts as fundamental as authorship, for example. Confronted with over 150 versions of the work, no two quite alike, what becomes of the concept of authorial control? And how can one assert with certainty which of the 150 or so versions is the 'correct' one, or even whether such a concept even makes sense in a pre-print culture.[202]

> Thus, the digitization of manuscripts and the creation of digital critical editions have not only provided new opportunities for textual criticism but also might even be viewed as enabling a

---

198 L. Berkes, in his chapter for this volume, asks: "should we expect online editions to conform to the norms of traditional printed editions or should we accept them as a slightly different form of publication?"

199 BABEU 2011, 36.

200 BODARD – GARCÉS 2009, 96.

201 BABEU 2011, 36.

202 NICHOLS 2009.

type of criticism that better respects the traditions of the texts or objects of analysis themselves[203].

Consider also the reflections of CAYLESS 2010 about the prominence of the transmission of content on its external appearance:

> [p]agination is a relatively fragile construct in the digital age", and textual "accretions" like commentaries, glosses and marginal notes, progressively gathered around the main text in its historical transmission, can be effectively encoded and represented in digital editions that not simply replicate print technologies.[204]

When we note (again after CAYLESS 2010, 162) that – functionally and theoretically – traditional commentary is a hypertext in print,[205] everything comes full circle, and it appears clearly how the new technologies can produce a very similar outcome as the ancient textual criticism described above. It can be argued, therefore, that a digital critical edition can develop into something completely different from the somehow 'old-fashioned' printed critical edition: namely, a further step in the fluid textual transmission of ancient sources.

6 Bibliography

ANDORLINI, I. (1981), *Ricette mediche nei papiri: note d'interpretazione e analisi di ingredienti (σμύρνα, καδμεία, ψιμύθιον)*, "Atti e Memorie dell'Accademia Toscana di Scienze e Lettere La Colombaria" 46, n.s. 32, 33–81 [repr. in *πολλὰ ἰατρῶν ἐστι συγγράμματα, I. Scritti sui papiri e la medicina antica*, a. c. di N. Reggiani, Firenze 2017, 37–48, and *II. Edizioni di papiri medici greci*, a. c. di N. Reggiani, Firenze 2018, forthcoming].

ANDORLINI, I. (1993), *L'apporto dei papiri alla conoscenza della scienza medica antica*, in *Aufstieg und Niedergang der römischen Welt*, II 37.1, hrsg. von W. Haase und H. Temporini, Berlin – New York, 458–562.

ANDORLINI, I. (1997a), *Progetto per il Corpus dei Papiri Greci di Medicina*, in *Akten des 21. Internationalen Papyrologenkongresses (Berlin 1995)*, hrsg. von B. Kramer, W. Luppe, H. Maehler und G. Poethke, Berlin – Boston, 17–24 [repr. in *πολλὰ ἰατρῶν ἐστι συγγράμματα, I. Scritti sui papiri e la medicina antica*, a. c. di N. Reggiani, Firenze 2017, 337–43].

ANDORLINI, I. (1997b), *Trattato o catechismo? La tecnica della flebotomia in PSI inv. CNR 85/86*, in *'Specimina' per il Corpus dei Papiri Greci di Medicina. Atti dell'incontro di studio (Firenze 1996)*, a. c. di I. Andorlini, Firenze, 153–68 [repr. in *πολλὰ ἰατρῶν ἐστι συγγράμματα, I. Scritti sui papiri e la medicina antica*, a. c. di N. Reggiani, Firenze 2017, 117–30].

---

203 BABEU 2011, 39.
204 CAYLESS 2010, 148. Quite interestingly, Cayless' picture exactly parallels the arguments brought by HANSON 1997 apropos of the transmission of ancient medical fragmentary texts, and the "accretive model of composition" (e.g. p. 305, see above) that she envisages to overcome the limits and rigidities of stemmatological interpretations.
205 Cf. also CAYLESS 2010, 170.

ANDORLINI, I. (1999), *Testi medici per la scuola: raccolte di definizioni e questionari nei papiri*, in *I testi medici greci. Tradizione e ecdotica. Atti del III Convegno internazionale (Napoli 1997)*, a c. di A. Garzya e J. Jouanna, Napoli, 7–15 [repr. in *πολλὰ ἰατρῶν ἐστι συγγράμματα, I. Scritti sui papiri e la medicina antica*, a c. di N. Reggiani, Firenze 2017, 286–93].

ANDORLINI, I. (2000), *Codici papiracei di medicina con scolî e commento*, in *Le commentaire entre tradition et innovation. Actes du colloque international de l'Institut des Traditions Textuelles (Paris et Villejuif 1999)*, éd. par M.-O. Goulet-Cazé, Paris, 37–52.

ANDORLINI, I. (2001), *Hippocrates, De fracturis 37 (PAberd 124r)*, in *Greek Medical Papyri I*, ed. by I. Andorlini, Firenze, 2–8 [repr. in *πολλὰ ἰατρῶν ἐστι συγγράμματα, II. Edizioni di papiri medici greci*, a c. di N. Reggiani, Firenze 2018, forthcoming].

ANDORLINI, I. (2003), *L'esegesi del libro tecnico: papiri di medicina con scolî e commento*, in *Papiri filosofici. Miscellanea di studi IV*, Firenze, 9–29 [repr. in *πολλὰ ἰατρῶν ἐστι συγγράμματα, I. Scritti sui papiri e la medicina antica*, a c. di N. Reggiani, Firenze 2017, 294–317].

ANDORLINI, I. (2006), *Il 'gergo' grafico ed espressivo della ricettazione medica antica*, in *Medicina e società nel mondo antico (Udine 2005)*, a c. di A. Marcone, Firenze, 142–67 [repr. in *πολλὰ ἰατρῶν ἐστι συγγράμματα, I. Scritti sui papiri e la medicina antica*, a c. di N. Reggiani, Firenze 2017, 15–36].

ANDORLINI, I. (2014), *Ippocratismo e medicina ellenistica in un trattato medico su papiro*, in *Hippocrate et les hippocratismes: médecine, religion, société. Actes du XIVe Colloque International Hippocratique (Paris 2012)*, éd. par J. Jouanna et M. Zink, Paris, 217–29 [repr. in *πολλὰ ἰατρῶν ἐστι συγγράμματα, I. Scritti sui papiri e la medicina antica*, a c. di N. Reggiani, Firenze 2017, 217–25].

ANDORLINI, I. (2017), *Il corpus dei papiri medici online: la piattaforma editoriale*, in *Atti del VII Colloquio Internazionale sull'Ecdotica dei testi medici greci (Procida 2013)*, a c. di A. Roselli, Napoli, in press [repr. in *πολλὰ ἰατρῶν ἐστι συγγράμματα, I. Scritti sui papiri e la medicina antica*, a c. di N. Reggiani, Firenze 2017, 380–90].

ANDORLINI, I. (2018), *From Prescription to Practice: The Evidence of two Medical Papyri from Roman Egypt*, in *Greek Medical Papyri: Text, Context, Hypertext. Proceedings of the DIGMEDTEXT International Conference (Parma 2016)*, ed. by N. Reggiani, Berlin – Boston, forthcoming.

ANDORLINI, I. – REGGIANI, N. (2012), *Edizione e ricostruzione digitale dei testi papiracei*, in *Diritto romano e scienze antichistiche nell'era digitale. Convegno di studio (Firenze 2011)*, a c. di N. Palazzolo, Torino, 131-46 [repr. in *πολλὰ ἰατρῶν ἐστι συγγράμματα, I. Scritti sui papiri e la medicina antica*, a c. di N. Reggiani, Firenze 2018, 363–75].

AST, R. (2017), *Signs of Learning in Greek Documents : The Case of* spiritus asper, in NOCCHI MACEDO – SCAPPATICCIO 2017, 143–58.

BABEU, A. (2011), *"Rome Wasn't Digitized in a Day": Building a Cyberinfrastructure for Digital Classics*, Washington (DC).

BAGNALL, R.S. (1995), *Reading Papyri, Writing Ancient History*, London - New York.

BAGNALL, R.S. (2012), *The* Amicitia Papyrologorum *in a Globalized World of Learning*, in *Actes du 26e Congrès International de Papyrologie (Genève 2010)*, éd. par P. Schubert, Genève, 1-5.

BAGNALL, R.S. – GAGOS, T. (2007), *The Advanced Papyrological Information System: Past, Present, and Future*, in *Proceedings of the 24th International Congress of Papyrology (Helsinki 2004)*, ed. by J. Frosen, T. Purola, and E. Salmenkivi, Helsinki, I, 59-74.

BARBIS LUPI, R. (1988), *La diplè obelismene: precisazioni terminologiche e formali*, in *Proceedings of the XVIII Inernational Congress of Papyrology (Athens 1986)*, ed. by B.G. Mandilaras, Athens, II, 473–6.

BARBIS LUPI, R. (1992), *Uso e forma dei segni di riempimento nei papiri letterari greci*, in *Proceedings of the XIXth International Congress of Papyrology (Cairo 1989)*, ed. by A.H.S. El-Mosalamy, Cairo, I, 503–10.

BARBIS LUPI, R. (1994), *La Paragraphos : analisi di un segno di lettura*, in *Proceedings of the 20th International Congress of Papyrologists (Copenhagen 1992)*, ed. by A. Bülow-Jacobsen, Copenhagen, 414–7.

BARNS, J.W.B. (1949), *Literary Texts from the Fayum*, CQ 43, 1–8.

BAUMAN, Z. (2000), *Liquid Modernity*, Malden (MA).

BAUMAN, Z. (2007), *Liquid Times. Living in an Age of Uncertainty*, Malden (MA).

BAUMAN, Z. (2011), *Culture in a Liquid Modern World*, Malden (MA).

BELL, H.I. (1953), *Abbreviations in Documentary Papyri*, in *Studies Presented to David Morre Robinson*, Saint Louis (MO), 424–33.

BERNARDELLI, A. (2000), *Intertestualità*, Firenze.

BERNARDELLI, A. (2010), *La rete intertestuale. Percorsi tra testi, discorsi e immagini*, Perugia.

BERTI, M. (2012), *L'intertestualità e la storiografia greca frammentaria*, in *Tradizione e trasmissione degli storici greci frammentari II. Atti del Terzo Workshop Internazionale (Roma 2011)*, a c. di V. Costa, Tivoli (RM), 439–58.

BERTI, M. (forthcoming), *Fragmentary Texts and Digital Libraries*, in *Philology in the Age of Corpus and Computational Linguistics*, ed. by G. Crane, A. Lüdeling, and M. Berti, CHS Publication, URL: http://www.fragmentarytexts.org/wp-content/uploads/2010/07/Berti-Fragmentary-Texts-and-Digiital-Libraries.pdf.

BERTONAZZI, F. (2018a), *Il lessico degli strumenti chirurgici nei papiri greci di medicina. Dalla digitalizzazione dei testi allo studio delle parole*, PhD Diss. Parma.

BERTONAZZI, F. (2018b), *Edizione digitale di P. Strasb. inv. 1187: un confronto tra il testo papiraceo e la tradizione manoscritta*, in *Proceedings of the 28th International Congress of Papyrology (Barcelona 2016)*, ed. by S. Torallas Tovar and A. Nodar, Barcelona, forthcoming.

BIRD, G.D. (2010), *Multitextuality in the Homeric Iliad. The Witness of the Ptolemaic Papyri*, Cambridge (MA) – London.

BLANCHARD, A. (1974), *Sigles et abbréviations dans les papyrus documentaires grecs. Recherches de paléographie*, London.

BODARD, G. – GARCÉS, J. (2009), *Open Source Critical Editions: A Rationale*, in *Text Editing, Print, and the Digital World*, ed. by M. Deegan and K. Sutherland, Farnham – Burlington (VT), 84‑98.

BOLTER, J.D. (1991), *The Computer, Hypertext, and Classical Studies*, AJP 112, 541‑5.

BONATI, I. (2015), *χύτρα*, in *Medicalia Online*, ed. by I. Andorlini, URL: http://www.papirologia.unipr.it/CPGM/medicalia/vocab/index.php?tema=141.

BONATI, I. (2016a), *Il lessico dei vasi e dei contenitori greci nei papiri.* Specimina *per un repertorio lessicale degli angionimi greci*, Berlin – New York.

BONATI, I. (2016b), *L'etichettatura del farmaco: radici antiche di una tradizione millenaria*, in *Medica-papyrologica. Miscellanea di studi presentati al Convegno Internazionale "Parlare la Medicina" (Parma 2016)*, a c. di N. Reggiani, Parma, 43–77.

BONATI, I. (2017), *L'uso della metafora nella microlingua greca della medicina*, in *La metafora e la sua traduzione: fra riflessioni teoriche e casi applicativi*, a c. di D. Astori, Parma, 83–100.

BONATI, I. (2018a), *Tra composti, suffissi e neologismi nella microlingua della medicina: alcuni* specimina *tratti dai papiri*, in *Parlare la medicina: fra lingue e culture, nello spazio e nel tempo. Atti del Convegno Internazionale (Parma 2016)*, a c. di N. Reggiani e F. Bertonazzi, Firenze, forthcoming.

BONATI, I. (2018b), *Medicalia Online: tecnicismi medici tra passato e presente*, in *Greek Medical Papyri: Text, Context, Hypertext. Proceedings of the DIGMEDTEXT International Conference (Parma 2016)*, ed. by N. Reggiani, Berlin – New York, forthcoming.

BONATI, I. (2018c), *La parola delle cose: nuove voci dal passato dei papiri*, in *Papiri, medicina antica e cultura materiale. Contributi in ricordo di Isabella Andorlini*, a c. di N. Reggiani, Parma, forthcoming.

BONATI, I. (2018d), *Between Text and Context: P.Oslo II 54 Revisited*, in *Proceedings of the 28th International Congress of Papyrology (Barcelona 2016)*, ed. by S. Torallas Tovar and A. Nodar, Barcelona, forthcoming.

BONATI, I. (2018e), *Definitions and Technical Terminology in the* Erôtapokriseis *on Papyrus*, in *"Where Does It Hurt?" Ancient Medicine in Questions and Answers*, ed. by M. Meeusen and E. Gielen, Leiden – Boston, forthcoming.

BONATI, I. – REGGIANI, N. (2018), *Mirrors in the Greek Papyri: Question of Words*, in *Mirrors and Mirroring: From Antiquity to the Early Modern Period. Proceedings of the Workshop (Wien 2017)*, ed. by L. Diamantopoulou and M. Gerolemou, forthcoming.

BOSCHETTI, F. (2007), *Methods to Extend Greek and Latin Corpora with Variants and Conjectures: Mapping Critical Apparatuses onto Reference Text*, in *Proceedings of the Corpus Linguistics Conference (Birmingham 2007)*, URL: http://ucrel.lancs.ac.uk/publications/CL2007/paper/150_Paper.pdf.

BRODERSEN, K. (1994), *Hippokrates und Artaxerxes. Zu P. Oxy. 1184v, P. Berol. inv. 7094v und 21137v + 6934v*, ZPE 102, 100–10.

CAYLESS, H.A. (2010), Ktema es aiei*: Digital Permanence from an Ancient Perspective*, in *Digital Research in the Study of Classical Antiquity*, ed. by G. Bodard and S. Mahony, Farnham ‐ Burlington (VT), 139‐50.

CLARYSSE, W. (1990), *Abbreviations and Lexicography*, AncSoc 21, 33–44.

COLOMO, D. (2017), *Quantity Marks in Greek Prose Texts on Papyrus*, in NOCCHI MACEDO – SCAPPATICCIO 2017, 97–126.

COLONNA, A. (1954), *Un antico commento ai* Theriaca *di Nicandro*, "Aegyptus" 34, 3–26.

CORAZZA, F. (2016), *New Recipes by Heras in P.Berol.Möller 13*, ZPE 198, 39–48.

CORAZZA, F. (2018a), *La digitalizzazione dei papiri medici di Antinoupolis*, in *Greek Medical Papyri: Text, Context, Hypertext. Proceedings of the DIGMEDTEXT International Conference (Parma 2016)*, ed. by N. Reggiani, Berlin – New York, forthcoming.

CORAZZA, F. (2018b), *The Antinoopolis Medical Papyri: A Case Study in Late Antique Medicine*, PhD Diss. Berlin.

CRIBIORE, R. (1996), *Writing, Teachers, and Students in Graeco-Roman Egypt*, Atlanta (GA).

CRIBIORE, R. (2018), *Genetic Criticism and the Papyri: Some Suggestions*, in *Greek Medical Papyri: Text, Context, Hypertext. Proceedings of the DIGMEDTEXT International Conference (Parma 2016)*, ed. by N. Reggiani, Berlin – New York, forthcoming.

DAMON, C. (2016), *Beyond Variants: Some Digital Desiderata for the Critical Apparatus of Ancient Greek and Latin Texts*, in *Digital Scholarly Editing. Theories and Practices*, ed. by M.J. Driscoll and E. Pierazzo, Cambridge (UK), 201‐18.

DEGANI, E. (1992), *Il mostro di Irvine*, "Eikasmòs" 3, 277‐8.

DEGNI, P. (1999), *Per uno studio sulle abbreviazioni greche. Dalle origini al IV secolo d.C.*, S&C 23, 63–74.

DEL MASTRO, G. (2017), *La ponctuation dans les papyrus grecs d'Herculaneum*, in *Signes dans les textes, textes sur les signes. Érudition, lecture et écriture dans le monde gréco-romain. Actes du colloque international (Liège 2013)*, éd. par G. Nocchi Macedo et M.C. Scappaticcio, Liège, 77–96.

DEPAUW, M. – STOLK, J. (2015), *Linguistic Variation in Greek Papyri: Towards a New Tool for Quantitative Study*, GRBS 55, 196‐220.

DICKEY, E. (2017), *Word Division in Bilingual Texts*, in NOCCHI MACEDO – SCAPPATICCIO 2017, 159–76.

DORANDI, T. (2014), *Ancient ἐκδόσεις: Further Lexical Observations on Some Galen's Texts*, "Lexicon Philosophicum" 2, 1–23, URL: http://lexicon.cnr.it/index.php/LP/article/view/398.

ESSLER, H. – RIAÑO RUFILANCHAS, D. (2016), *'Aristarchus X' and Philodemus: Digital Linguistic Analysis of a Herculanean Text Corpus*, in *Proceedings of the 27th International Congress of Papyrology (Warsaw 2013)*, ed. by J. Urbanik, T. Derda, and A. Łajtar, Warsaw, I, 491‐501.

FINGLASS, P.J. (2015), *Reperformances and the Transmission of Texts*, "Trends in Classics" 7, 259–76.

FAUSTI, D. (1989), *P.Strasb. inv. gr. 1187: testo chirurgico (Eliodoro?)*, "Annali della Facoltà di Lettere e Filosofia, Firenze" 10, 157–69.

Fischer, K.-D. (1998), *Beiträge zu den pseudosoranischen* Quaestiones medicinales, in *Text and Tradition. Studies in Ancient Medicine and its Transmission*, eds. K.-D. Fischer, D. Nickel, and P. Potter, Leiden – Boston – Köln, 1–41.

Funari, R. (2017), *Segni di interpunzione e di lettura nei frammenti storici latini da papiro e pergamena rinvenuti nell'Egitto*, in Nocchi Macedo – Scappaticcio 2017, 177–202.

Gagos, T. (2001), *The University of Michigan Papyrus Collection: Current Trends and Future Perspectives*, in *Atti del XXII Congresso Internazionale di Papirologia (Firenze 1998)*, a c. di I. Andorlini, G. Bastianini, M. Manfredi e G. Menci, Firenze, II, 511‑37.

Garofalo, I. – Fortuna, S. – Lami, A. – Roselli, A. (2010), curr., *Sulla tradizione indiretta dei testi medici greci: le traduzioni. Atti del III seminario internazionale (Siena 2009)*, Pisa – Roma.

Genette, G. (1992), *The Architext: An Introduction*, Berkeley (CA) [*Introduction à l'architexte*, Paris 1979].

Genette, G. (1997), *Palimpsests: Literature in the Second Degree*, Lincoln (OR) [*Palimpsestes: la littérature au second degré*, Paris 1982].

Gignac, F.T. (1976), *A Grammar of the Greek Papyri of the Roman and Byzantine Periods, I: Phonology*, Milano.

Gonis, N. (1009), *Abbreviations and Symbols*, in *The Oxford Handbook of Papyrology*, ed. by R.S. Bagnall, Oxford, 170–8.

Hanson, A.E. (1970), *P. Antinoopolis 184: Hippocrates,* Diseases of Women, in *Proceedings of the Twelfth International Congress of Papyrology*, ed. by D.H. Samuel, Toronto, 213–22.

Hanson, A.E. (1985), *Papyri of Medical Content*, YCS 28, 25–47.

Hanson, A.E. (1996), *Introduction*, in Youtie 1996, xv–xxv.

Hanson, A.E. (1997), *Fragmentation and the Greek Medical Writers*, in *Collecting Fragments / Fragmente Sammeln*, ed. by G.W. Most, Göttingen, 289–314

Hanson, A.E. (1998), *Galen: Author and Critic*, in *Editing Texts / Texte edieren*, ed. by G.W. Most, Göttingen, 22–52.

Hanson, A.E. (2002), *Papyrology: A Discipline in Flux*, in *Disciplining Classics / Altertumswissenschaft als Beruf*, ed. by G.W. Most, Göttingen, 191‑206.

Hanson, A.E. (2010), *Doctors' Literacy and Papyri of Medical Content*, in *Hippocrates and Medical Education*, ed. by M. Horstmanshoff, Leiden – Boston, 187–204.

Hanson, A.E. – Mattern, S.P. (2001), *Medical Catechism*, in *Greek Medical Papyri I*, ed. by I. Andorlini, Firenze, 71–83.

Jördens, A. (2017), *Entwurf und Reinschrift – oder: Wie bitte ich um Entlassung aus der Untersuchungshaft*, "Chiron" 47, 271–302.

Kollesch, J. (1963), *Untersuchungen zu den pseudogalenischen* Definitiones Medicae, Berlin.

Lamé, M. (2014), *Primary Sources of Information, Digitization Processes and Dispositive Analysis*, in *Proceedings of the Third AIUCD Annual Conference on Humanities and Their Methods in the Digital Ecosystem (Bologna 2014)*, URL: https://doi.org/10.1145/2802612.2802645.

Leith, D. (2007), *A Medical Treatise "On Remedies"? P.Turner 14 Revised*, BASP 44, 125–35.

Lüdeling, A. (2011), *Corpora in Linguistics: Sampling and Annotation*, in *Going Digital. Evolutionary and Revolutionary Aspects of Digitization*, ed. by K. Grandin, New York, 220‑43.

Lüdeling, A. ‑ Kytö, M. (2008–09), eds., *Corpus Linguistics. An International Handbook*, I‑II, Berlin.

Luiselli, R. (2010), *Authorial Revision of Linguistic Style in Greek Papyrus Letters and Petitions (AD IIV)*, in *The Language of the Papyri*, ed. by T.V. Evans and D.D. Obbink, Oxford, 88‑94.

Macé, C. – Baret, P. – Bozzi, A. – Cignoni, L. (2006), eds., *The Evolution of Texts: Confronting Stemmatological and Genetical Methods. Proceedings of the International Workshop (Louvain-la-Neuve 2004)*, Pisa – Roma.

Magdelaine, C. (2004), *Un nouveau questionnaire ophtalmologique (PStrasb gr. inv. 849)*, in *Testi medici su papiro. Atti del Seminario di studio (Firenze 2002)*, a c. di I. Andorlini, Firenze, 63–77.

MAGNANI, M. (2008), *"Sapere ex indicibus"*, in *Scienze umane e cultura digitale. Atti della XVI Settimana della Cultura Scientifica (Parma 2006)*, Fiesole (FI), 127‑38.

MANETTI, D. – ROSELLI, A. (1994), *Galeno commentatore di Ippocrate*, in *Aufstieg und Niedergang der römischen Welt*, II 37/2, hrsg. von H. Temporini, Berlin, 1557–65.

MANIACI, M. (2002), *"La serva padrona". Interazioni fra testo e glossa nella pagina del manoscritto* in *Talking to the Text: Marginalia from Papyri to Print. Proceedings of the Conference (Erice 1998)*, ed. by V. Fera, G. Ferraù, and S. Rizzo, Messina, I, 3–35.

MARAVELA, A. – LEITH, D. (2007), *A Medical Catechism on Tumours from the Collection of the Oslo University Library*, in *Proceedings of the 24th International Congress of Papyrology (Helsinki 2004)*, ed. by J. Frösén, T. Purola, and E. Salmenkivi, Helsinki, 637–50.

MARAVELA, A. – REGGIANI, N. (2018), *Medical Scribes at Work: Exploring Linguistic Variation in Greek Medical Papyri*, in *Act of the Scribe: Interfaces between Scribal Work and Language Use. Proceedings of the Workshop (Athens 2017)*, ed. by S. Dahlgren, H. Halla-aho, M. Leiwo, M. Vierros, Helsinki, forthcoming.

MARGANNE, M.-H. (1980), *Une étape dans la transmission d'une préscription médicale: P. Berl. Möller 13*, in *Miscellanea papyrologica*, a c. di R. Pintaudi, Firenze, 179–83.

MARGANNE, M.-H. (1981a), *Inventaire analytique des papyrus grecs de medecine*, Genève.

MARGANNE, M.-H. (1981b), *Un fragment du medecin Herodote: P. Tebt. II 272*, in *Proceedings of the XVI Int. Congr. of Papyrology (New York 1980)*, Chico, 73–8.

MARGANNE, M.-H. (1986), *La 'Cirurgia Eliodori' et le P. Genève inv. 111*, "Etudes de Lettres", 65–73.

MARGANNE, M.-H. (1998), *La chirurgie dans l'Égypte gréco-romaine d'après les papyrus littéraires grecs*, Leiden – Boston – Köln.

MARGANNE, M.-H. (2004), *Le livre medical dans le monde gréco-romain*, Liège.

MARGANNE, M.-H. – MERTENS, P. (1997), *Medici et Medica. 2e edition (Etat au 15 janvier 1997 du fichier MP3 pour les papyrus medicaux litteraires)*, in *'Specimina' per il Corpus dei Papiri Greci di Medicina. Atti dell'Incontro di studio (Firenze 1996)*, a c. di I. Andorlini, Firenze, 3‑71.

MCNAMEE, K. (1981), *Abbreviations in Greek Literary Papyri and Ostraca*, Chico (CA).

MCNAMEE, K. (1985), *Abbreviations in Greek Literary Papyri and Ostraca: Supplement, with List of Ghost Abbreviations*, BASP 22, 205–25.

MCNAMEE, K. (2007), *Annotations in Greek and Latin Texts from Egypt*, Oakville (CT)

MCNAMEE, K. (2017), *Sigla in Late Greek Literary Papyri*, in NOCCHI MACEDO – SCAPPATICCIO 2017, 127–42.

MESSERI SAVORELLI, G. – PINTAUDI, R. (2002), *I lettori dei papiri: dal commento autonomo agli scolii*, in *Talking to the Text: Marginalia from Papyri to Print. Proceedings of the Conference (Erice 1998)*, ed. by V. Fera, G. Ferraù, and S. Rizzo, Messina, I, 37–57.

NIEDDU, G.F. (1992), *Il ginnasio e la scuola: scrittura e mimesi del parlato*, in *Lo spazio letterario della Grecia antica*, a c. di G. Cambiano, L. Canfora e D. Lanza, Roma, I.1, 555–85.

NIELSEN, E. (2000), *A Catalog of Duplicate Papyri*, ZPE 129, 187–214.

NOCCHI MACEDO, G. – SCAPPATICCIO, M.C. (2017), édd., *Signes dans les textes, textes sur les signes. Érudition, lecture et écriture dans le monde gréco-romain. Actes du colloque international (Liège 2013)*, Liège.

NODAR DOMÍNGUEZ, A. (2017), *Los signos de lectura más antiguos en papiro*, in NOCCHI MACEDO – SCAPPATICCIO 2017, 61–76.

NUTTON, V. (1972), *Galen and Medical Autobiography*, PCPS 18, 50–62.

NUTTON, V. (1990), *The Patient's Choice: A New Treatise by Galen*, CQ 40, 236–57.

OWENS, T. (2011), *Defining Data for Humanists: Text, Artifact, Information or Evidence?*, "Journal of Digital Humanities" 1.1, URL: http://journalofdigitalhumanities.org/1-1/defining-data-forhumanists-by-trevor-owens.

PASSAROTTI, M. (2006), *Towards Textual Drift Modelling in Computational Philology*, in MACÉ – BARET – BOZZI – CIGNONI 2006, 63‑88.

Petit, C. (2009), *Galien: Le médecin. Introduction*, Paris.

Pierazzo, E. (2008), *Digital Genetic Editions: The Encoding of Time in Manuscript Transcription*, in *Text Editing, Print and the Digital World. Digital Research in the Arts and Humanities*, ed. by M. Deegan and K. Sutherland, Aldershot, 169–86.

Polacco, M. (1998), *L'intertestualità*, Roma – Bari.

Porter, S. – O'Donnell, M.B. (2010), *Building and Examining Linguistic Phenomena in a Corpus of Representative Papyri*, in *The Language of the Papyri*, ed. by T.V. Evans and D.D. Obbink, Oxford 2010, 287–311.

Reggiani, N. (2015), *A Corpus of Literary Papyri Online: the Pilot Project of the Medical Texts via So-SOL*, in *Antike Lebenswelten Althistorische und papyrologische Studien*, hrsg. von R. Lafer und K. Strobel, Berlin – New York, 341–52.

Reggiani, N. (2016a), *The Corpus of Greek Medical Papyri and Digital Papyrology: New Perspectives from an Ongoing Project*, in *Altertumswissenschaften in a Digital Age: Egyptology, Papyrology and beyond. Proceedings of a Conference and Workshop in Leipzig (November 2015)*, ed. by M. Berti and F. Naether, Leipzig, URL: http://nbn-resolving.de/urn:nbn:de:bsz:15-qucosa-201726.

Reggiani, N. (2016b), *Catechism,* in *Medicalia Online*, ed. I. Andorlini, URL: http://www.papirologia.unipr.it/CPGM/medicalia/vocab/index.php?tema=8.

Reggiani, N. (2016c), *Data Processing and State Management in Late Ptolemaic and Roman Egypt: The Project 'Synopsis' and the Archive of Menches*, in *Proceedings of the 27th International Congress of Papyrology (Warsaw 2013)*, ed. by. T. Derda, A. Łajtar, and J. Urbanik, Warsaw, III, 1415–44.

Reggiani, N. (2017), *Digital Papyrology I. Tools, Methods and Trends*, Berlin – New York.

Reggiani, N. (2018a), *Linguistic and Philological Variants in the Papyri: A Reconsideration in Light of the Digitization of the Greek Medical Papyri*, in *Greek Medical Papyri – Text, Context, Hypertext. Proceedings of the DIGMEDTEXT International Conference (Parma 2016)*, ed. by N. Reggiani, Berlin – New York, forthcoming.

Reggiani, N. (2018b), *The Corpus of Greek Medical Papyri Online and the Digital Edition of Ancient Documents*, in *Proceedings of the 28th International Congress of Papyrology (Barcelona 2016)*, ed. by S. Torallas Tovar and A. Nodar, Barcelona, forthcoming.

Reggiani, N. (2018c), *The Digital Edition of Ancient Sources as a Further Step in the Textual Transmission*, in *Proceedings of the Workshop "Digital Classics III: Re-thinking Text Analysis" (Heidelberg 2017)*, ed. by A. Novokhatko, S. Chronopoulos, and F.K. Maier, forthcoming.

Reggiani, N. (2018d), *Isabella Andorlini, la papirologia medica e il progetto DIGMEDTEXT*, in *Papiri, medicina antica e cultura materiale. Contributi in ricordo di Isabella Andorlini*, a cura di N. Reggiani, Parma, forthcoming.

Reggiani, N. (2018e), *Ancient Doctors' Literacies and the Digital Edition of Papyri of Medical Content*, "Classics@", forthcoming.

Reggiani, N. (2018f), *Ispezioni e perizie ufficiali nell'Egitto romano: il* corpus *dei rapporti professionali (*prosphōnēseis*)*, in *Lavoro manuale e lavoro intellettuale nella società romana*, a c. di A. Marcone e P. Porena, Roma, forthcoming.

Reggiani, N. (2018g), *Transmission of Recipes and Receptaria in Greek Medical Writings on Papyrus Between Ancient Text Production and Modern Digital Representation*, in *"Cupis volitare per auras": Books, Libraries and Textual Transmission from the Ancient to the Medieval World. Proceedings of the First Postgraduate Conference (Bari 2016)*, ed. by E. Barile, R. Berardi, N. Bruno, M. Filosa, and L. Fizzarotti, forthcoming.

Reggiani, N. (2018h), *Digitizing Medical Papyri in Question-and-Answer Format*, in *"Where Does It Hurt?" Ancient Medicine in Questions and Answers*, ed. by E. Gielen and M. Meeusen, Leiden – Boston, forthcoming.

Reggiani, N. (2018i), *Herbal*, in *Medicalia Online*, ed. I. Andorlini, URL: http://www.papirologia.unipr.it/CPGM/medicalia/vocab/index.php?tema=11, forthcoming.

REGGIANI, N. (2018j), *Prescrizioni mediche e supporti materiali nell'Antichità*, in *Parlare la medicina: fra lingue e culture, nello spazio e nel tempo. Atti del Convegno Internazionale (Parma 2016)*, a cura di N. Reggiani e F. Bertonazzi, Firenze, forthcoming.

RIAÑO RUFILANCHAS, D. (2014), *The "Grammatically Annotated Philodemus" Project*, CErc 44, 155‒65.

ROMANELLO, M. – BERTI, M. – BOSCHETTI, F. – BABEU, A. – CRANE, G. (2009), *Rethinking Critical Editions of Fragmentary Texts by Ontologies*, in *Rethinking Electronic Publishing: Innovation in Communication Paradigms and Technologies. Proceedings of 13th International Conference on Electronic Publishing (Milano 2009)*, Milano, 155‒74, URL: https://elpub.architexturez.net/doc/oai-elpub-id-158-elpub2009.

ROUED-CUNLIFFE, H. (2009), *Textual Analysis Using XML: Understanding Ancient Textual Corpora*, in *5th IEEE Conference on e-Science 2009*, URL: http://esad.classics.ox.ac.uk/IEEE_Paper_HRa19b.pdf_%3b%20modification-date%3d_Thu%2c%2007%20Jul%202011%2013_19_56%20%2b0000_%3b%20size%3d1625627%3b?option=com_docman&task=doc_download&gid=30&Itemid=78.

SAVIGNAGO, L. (2008), *Eisthesis. Il sistema dei margini nei papiri dei poeti tragici*, Alessandria.

SCHUBERT, P. (2009), *Editing a Papyrus*, in *The Oxford Handbook of Papyrology*, ed. by R.S. Bagnall, Oxford, 197‒215.

SINCLAIR, J. (1996), *EAGLES [Expert Advisory Group on Language Engineering Standards]: Preliminary Recommendations on Corpus Typology*, URL: http://www.ilc.cnr.it/EAGLES96/corpustyp/corpustyp.html.

STOLK, J. (2015a), *Case Variation in Greek Papyri. Retracing dative case syncretism in the language of the Greek documentary papyri and ostraca from Egypt (300 BCE – 800 CE)*, PhD Diss. Oslo.

STOLK, J. (2015b), *Dative by Genitive Replacement in the Greek Language of the Papyri: A Diachronic Account of Case Semantics*, "Journal of Greek Linguistics" 15, 91–121.

STOOP, J. (2014), *Two Copies of a Royal Petition from Kerkeosiris*, ZPE 189, 185–93.

TOMASI, F. – ZAJA, P. (2002), *Proposte per un'edizione ipertestuale di postillati cinquecenteschi*, in *Talking to the Text: Marginalia from Papyri to Print. Proceedings of the Conference (Erice 1998)*, ed. by V. Fera, G. Ferraù, and S. Rizzo, Messina, II, 721–52.

TOMSIN, A. (1970), *Les papyrologues et le travail papyrologique par ordinateur*, in *Proceedings of the Twelfth International Congress of Papyrology*, ed. by D.H. Samuel, Toronto, 471‒84.

TOTELIN, L.M.V. (2009a), *Hippocratic Recipes. Oral and Written Transmission of Pharmacological Knowledge in Fifth- and Fourth-Century Greece*, Leiden – Boston.

TOTELIN, L.M.V. (2009b), *Galen's Use of Multiple Manuscript Copies in His Pharmacological Treatises*, in *Authorial Voices in Greco-Roman Technical Writing*, eds. L. Taub and A. Doody, Trier, 81–92.

TURNER, E.G. (1987), *Greek Manuscripts of the Ancient World. Second Edition Revised and Enlarged*, ed. by P.J. Parsons, London [1970].

VANNINI, L. (2012), *Papiri con edizioni commentate*, in *Actes du 26e Congrès International de Papyrologie (Genève 2010),* éd. par P. Schubert, Genève, 801–5.

VÉRONIS, J. (2000), ed., *Parallel Text Processing: Alignment and Use of Translation Corpora*, Dordrecht.

WILLIS, W.H. (1984), *The Duke Data Bank of Documentary Papyri*, in *Atti del XVII Congresso Internazionale di Papirologia (Napoli)*, Napoli, I, 167‒73.

WORTON, M. – STILL, J. (1990), *Intertextuality. Theories and Practices*, Manchester – New York.

YOUTIE, H.C. (1963), *The Papyrologist: Artificer of Fact*, GRBS 4, 19–32.

YOUTIE, L.C. (1996), *The Michigan Medical Codex*, ed. by A.E. Hanson, Atlanta (GA) [= *The Michigan Medical Codex*, 1986–87, ZPE 65, 123–49; 66, 149–56; 67, 83–95; 69, 163–9; 70, 73–103].

YUEN-COLLINGRIDGE, R. – CHOAT, M. (2012), *The Copyist at Work: Scribal Practice in Duplicate Documents*, in *Actes du 26e Congres international de papyrologie (Genève 2010)*, éd. par P. Schubert, Genève, 827–34.

Rodney Ast, Holger Essler

Anagnosis, Herculaneum, and the *Digital Corpus of Literary Papyri*

1 Overview

The *Digital Corpus of Literary Papyri* (DCLP) initiative was launched in July 2013 with support from the NEH/DFG Bilateral Digital Programme and upon successful completion of a one-year planning grant under the same program.[1] The project, which ended in August 2017, built the necessary framework for a large-scale corpus of literary papyri on the basis of infrastructure already in place at www.papyri.info for documentary papyri. And by "literary papyri" we mean both canonical literary genres, such as epic, lyric, drama, oratory, etc., and so-called para- or subliterary texts, whether magical, medical, or school. All content has been encoded in the well-established XML-TEI format known as EpiDoc. An attempt has been made to customize search, browse, and editing functionality to the needs of individuals who deal with literary texts. At the same time we have wanted to encourage engagement with all extant papyrological sources, both literary and documentary. As a result, users can search across the entire corpus of texts in the *Duke Databank of Documentary Papyri* (DDbDP) and DCLP.

Despite being headquartered in New York and Heidelberg, the DCLP has profited from collaboration with a large number of researchers and institutions across Europe and the United States. Mark Depauw and Willy Clarysse in Leuven shared metadata belonging to the *Leuven Database of Ancient Books* (LDAB) for nearly 15,000 objects. This information constitutes the backbone of each DCLP record (Fig. 1 with metadata from LDAB).

1 Principal investigators on the project were Rodney Ast (DFG) at the University of Heidelberg and Roger Bagnall (NEH) at New York University. Both Ast and Holger Essler have presented the project on numerous occasions, and a description of the initiative can be found also in Reggiani 2017, 250–4.

Browse: DDbDP HGV APIS DCLP Authors TM Number or Search: Data Bibliography

P.Ryl. 3 478 = Trismegistos 62954 = LDAB 4146

metadata☑ text☑ open in editor Canonical URI: http://lдаb.info/dclp/62954
TM data☑ DCLP data☑ transcription☑

DCLP/LDAB Data [xml]

Title	P. Ryl. 3 478
Work	Vergilius, Aeneis
Content	Vergilius; Aeneis 01.235-243, 247-261, 270-274, 406-414, 418-426, 633-640, 702-707, 711-719 with literal Greek translation - school book
Principal Edition	P.Ryl. 3 478
Fragments	Manchester, John Rylands Library Gr. 478; Milan, Università Cattolica P. Med. number unknown; Alexandria, Bibliotheca Alexandrina (exhibited) [85644]
Support Material	papyrus
Date	300 - 399 [More from the period between 300 CE and 399 CE]
Origin	Found: Oxyrynchos (U, Egypt); written: Egypt [More from Found: Oxyrynchos (U, Egypt); written: Egypt]
Form and Layout	papyrus codex (5 fol.) (columns: 2, written lines: [42 ca.], pagination: 0)
Script Type	early half-uncial
Genre	poetry; epic
Culture	literature
Religion	classical
Images	enriqueta.man.ac.uk/.../341; ipap.csad.ox.ac.uk/.../478; enriqueta.man.ac.uk/.../341
Print Illustrations	CLA 02 + 010 227; JJP 4 (1950), pl.III; P.Med. 1 pl.I-II = CLA 03 367; ZPE 11, pl.IX-XVI; R. Seider, Paläographie Lateinischer Papyri II.1 no. 38 pl. XXI; G. Cavallo, Ricerche sulla maiuscola biblica (1967), pl.33; B.Rochette, Le latin dans le monde grec, pl.II; G. Cavallo, La scrittura greca e latina dei papiri, 2008, p. 171 [143]
Availability	
Origin	[More from Found: Oxyrynchos (U, Egypt); written: Egypt]

Fig. 1: LDAB metadata.

Similarly, cooperation with Duke University's Duke Collaboratory for Classics Computing (DC3) allowed the DCLP team to build on the standards and tools in place at *Papyri.info*. Holger Essler and his team in Würzburg, together with Gianluca Del Mastro in Naples and Daniel Riaño in Madrid, were responsible for the addition of hundreds of files containing bibliography, links to sketches, engravings and photographs, and transcriptions of Herculaneum papyri. The Parma Medical Project, under the leadership of Isabella Andorlini† and Nicola Reggiani, produced extended editions of over 200 medical papyri. Furthermore the project benefited from the participation of numerous students and scholars who have both entered transcriptions and vetted submissions.

Before detailing what an individual DCLP record might contain and what the project seeks to cover over the long term, we will first speak more generally about the nature and scope of the initiative. The DCLP is not the first large-scale online resource for classical literature. The *Thesaurus Linguae Graecae* and the *Perseus Digital Library* both offer searchable and browsable transcriptions of ancient texts.[2] The most obvious difference with these initiatives is the fact that DCLP covers only papyrological evidence, including papyri, pre-medieval parchment, *ostraka*, tablets, dipinti, and other non-inscriptional evidence. It does not incorporate medieval manuscripts, and its interest is not strictly in the 'text' *per se*, but rather in the text as it appears on any given material substrate. In this respect, it focuses on the inscribed object as a whole, and tries to account for extra-textual elements such as layout and non-verbal signs

---

2 The former, which is largely confined to Greek literature, is hosted at http://stephanus.tlg.uci.edu; the latter, which comprises both Greek and Latin, can be found at http://www.perseus.tufts.edu/hopper. Colleagues at Tufts University and the University of Leipzig are working in a much more ambitious project called the *Open Greek and Latin Project* (OGL) to expand *Perseus* in order to include all Greek and Latin texts, http://www.dh.uni-leipzig.de/wo/projects/open-greek-and-latin-project. DCLP has the potential both to enhance OGL and benefit from it.

(e.g., *paragraphoi*, *diplai*, etc.), in addition to the written words.[3] As a result, a higher premium is placed on accurate decipherment of all elements of the textual witness. This is not to say that a photographic representation of these elements is offered in the HTML text, but links to photographs are provided when possible, so that the user can observe all features of the inscribed object.

We have also tried to make it easy to discover content in the DCLP. Text-search functionality is the same as in *Papyri.info*, as are many of the browsing options, and we have retained the faceted browsing capability. In addition to finding texts by their publication numbers, provided the publications are known to the *Checklist of Editions* (http://www.papyri.info/docs/checklist), one can also locate them by TM numbers (Fig. 2 with three browsing options).

Fig. 2: DCLP browsing options.

The latter is the most effective means of finding a known object, since every item has a TM number, but prior knowledge is a prerequisite, and in that respect it hardly can be described as a true browsing function. The author/works browsing option, on the other hand, represents an entry point that should be welcomed by the literature-minded user who wants to see how many papyri by a known author survive (Fig. 3). In addition to giving the names of extant works by known authors, the system also says when there is a Greek text available. Here again, though, DCLP is not charitable to ignorance: works by unknown authors do not appear in the list.

---

3 The project is very much in tune with current efforts in classical studies to account for physical aspects of ancient witnesses. One example of these efforts is the University of Heidelberg's Sonder-forschungsbereich 933, *Materiale Textkulturen*; more information about this SFB can be found at http://www.materiale-textkulturen.de.

Euripides
- Aigeus
  - Euripides, Aigeus 59946 (na;;59946) (Greek text available) view in dclpxsltbox | github
- Alcestis
  - Euripides, Alcestis 68623 (p.oxy;67;4546) (Greek text available) view in dclpxsltbox | github
  - Euripides, Alcestis 68624 (p.oxy;67;4547) (Greek text available) view in dclpxsltbox | github
- Alcmaeon
  - Euripides, Alcmaeon 59812 (psi;13;1302) view in dclpxsltbox | github
- Alexandros
  - Euripides, Alexandros 115700 (na;;115700) view in dclpxsltbox | github
- Andromache
  - Euripides, Andromache 59873 (p.oxy;67;4551) (Greek text available) view in dclpxsltbox | github
  - Euripides, Andromache 59824 (p.oxy;31;2543) view in dclpxsltbox | github
  - Euripides, Andromache 59857 (p.harris;1;39) view in dclpxsltbox | github
  - Euripides, Andromache 59878 (na;;59878) view in dclpxsltbox | github
  - Euripides, Andromache 59833 (p.oxy;22;2335) view in dclpxsltbox | github
  - Euripides, Andromache 59897 (na;;59897) view in dclpxsltbox | github
  - Euripides, Andromache 59931 (p.hib;1;25) view in dclpxsltbox | github
  - Euripides, Andromache 59899 (p.ross.georg;1;8) view in dclpxsltbox | github
  - Euripides, Andromache 59858 (p.oxy;3;449) view in dclpxsltbox | github
  - Euripides, Andromache 59876 (na;;59876) view in dclpxsltbox | github
  - Euripides, Andromache 68625 (p.oxy;67;4552) (Greek text available) view in dclpxsltbox | github
  - Euripides, Andromache 68627 (p.oxy;67;4554) (Greek text available) view in dclpxsltbox | github
  - Euripides, Andromache 68626 (p.oxy;67;4553) (Greek text available) view in dclpxsltbox | github
  - Euripides, Andromache 68628 (p.oxy;67;4555) (Greek text available) view in dclpxsltbox | github

Fig. 3: DCLP authors and works.

DCLP is not only a portal to information about published papyri, it is also a curatorial platform. Using the same editor employed at *Papyri.info*, the system allows users to propose emendations to published texts and to contribute introductions, commentaries, translations, as well as updated, machine-readable bibliography. All new content is vetted by members of the editorial board. The DCLP instance of Berlin P. 12310, an *ostrakon* preserving five verses of Theognis (vv. 434–8) plus an unidentified comedy fragment, is an example of an emended text.[4] The Theognis verses were thought by the original editor to be inferior to those preserved in other witnesses, one of which is Plato's *Meno* (95e). Even though the first lines are transposed and thus differ from the Theognidean manuscript tradition, the text is not as banal as the original editor thought. The ordering of the lines on the *ostrakon* agrees with the text found in Plato and the unique and incomprehensible reading of πάλιν that is printed by the first ed-

---

4 See http://litpap.info/dclp/62823, which contains extensive bibliography.

itor of the *ostrakon* in line 3 has turned out upon closer examination by Julia Lougovaya to be a mistaken reading.[5] The sherd actually has πολλούς, which is attested in all other witnesses. Lougovaya has emended the online text accordingly (Fig. 4, http://litpap.info/dclp/62823).

Fig. 4: DCLP text, apparatus, and notes.

The Parma medical project has been a significant contributor of new content, including extended editions of many papyri. P.Yale II 134 is representative of this work.[6] The introduction to the online edition, which is authored by N. Reggiani, briefly describes the content of the text, which is a set of iatromagical prescriptions. This is followed by the edition, critical apparatus, and commentary. The header gives bibliographical references and information about the date, provenance, physical dimensions of the fragments, etc. It also contains a link to an image of the papyrus at the host institution. While mostly based on previous printed editions, the DCLP record is itself a stand-alone edition.

---

5 The *ostrakon* was first published by Viereck 1925, 254–5, and subsequently in Pordomingo 2013, no. 27. Wolfgang Luppe arrived independently at the same conclusion about Viereck's reading of πάλιν in his re-edition of the *ostrakon* in *CPF* II, 2, pp. 358–9.

6 See http://www.litpap.info/dclp/64529.

2 Herculaneum

The texts of the Herculaneum papyri were also encoded on the basis of a reference edition, but provide additional bibliography and links to images. Our starting point was the *Thesaurus Herculanensium Voluminum*, the first full-text database of the Herculaneum papyri. It currently contains 26 texts with some 20k lines that were entered jointly by the Centro Internazionale per lo Studio dei Papiri Ercolanesi (CISPE) and the Würzburg Institute of Classics.[7] Since we firmly believe that uniting the Herculaneum papyri with the literary papyri on a single platform is an important step towards the unity of our discipline and, in particular, towards the development of Herculaneum Papyrology, all texts were transferred to DCLP and further work was and will be based on this database.[8] Currently DCLP comprises 117 editions of Herculaneum papyri and can be considered fairly complete for this group.

Since new editions of Herculaneum papyri tend to take decades, it was necessary to refrain from reediting or rechecking readings on the original papyri. In order to make the texts available as soon as possible we had to compromise even further: text entry was strictly based on the most recent comprehensive editions. Especially where readings are highly disputed, it would have been impossible to decide without thorough study of the whole papyrus. Instead, we were striving to provide a complete bibliography of editions of Herculaneum papyri. Currently this first-ever complete index comprises more than 1,400 records, of which 387 are reference editions, the central focus and basis of the digital text. In general, there are four categories (see the metadata in Fig. 5).

a) Reference editions: the bases followed for text entry
b) Previous editions: any edition predating the reference edition, which may be consulted in the future for the apparatus and concordances
c) Partial editions published after the publication of the reference edition
d) Readings: individual readings without establishing a syntactic connection, published after the reference edition

Although many partial editions provide substantial improvements to the text, their incorporation was not feasible at this stage. Besides, the current software does not allow to reference parts of text to particular editions.

The prerequisite for a uniform data structure was the establishment of the relationship between the canonical numbers of *Trismegistos* (http://www.trismegistos.org) and the LDAB and the traditional inventory numbers of the Herculaneum papyri (there referred to as Gigante Numbers, following the Catalogue edited by him).[9]

---

7 (12 November 2017).

8 A list of Herculaneum texts entered is available at: http://epikur-wuerzburg.de/thv.

9 GIGANTE 1979.

While Trismegistos and LDAB assign an identification number to each ancient book or scroll, the inventory numbers of the Herculaneum papyri refer to the fragments as they were inventoried in the 18th century. Several scrolls were already broken in pieces at the time of the excavation, some were cut to facilitate unrolling, and every piece, be it a stack of several layers from the outer part of the scroll or a part from the central cylinder, was assigned a different inventory number. Over the years, these pieces had a story of their own, and might now be in very different condition. In addition, difficulty in discerning the text on the darkened surface also resulted in miscataloguing single fragments. Thus there are several cases where a single inventory number contains fragments belonging to different papyri while often a single scroll (corresponding to one TM number) has to be reconstructed from several inventory numbers.[10] In several cases fragments that have been edited separately are now proven to belong to the same scroll. An example is TM 62499, containing Philodemus *De rhetorica II* (cf. Fig. 5). The scroll can be reconstructed from five inventory numbers, but no comprehensive edition is available. We have thus limited ourselves to assembling the text from the four reference editions that cover a maximum of the text preserved. Since they were published by different editors at different times (from 1892 to 1977), it is not surprising they differ considerably in scope and method.

Fig. 5: Metadata of TM 62499, Philodemus *On rhetoric* 2.

---

10 Examples are TM 62400, Philodemus *De pietate*, and TM 62419, Philodemus *De poematis* 1 (cf. e.g. the scheme in OBBINK 1996, 43; JANKO 2003, 105).

As a consequence of their preservation and the combination of fragments, the surviving remains of a Herculaneum papyrus are on average considerably larger than those of the Egyptian fragments. They normally exceed 600 lines and arrive at a maximum of nearly 4,500 lines for Philodemus *De musica IV*, distributed among more than one hundred fragments. Since for many fragments several images are available online, the traditional separation of text and metadata as respected in the display of *Papyri.info* resulted in extremely long lists of information without any visible connection to the corresponding texts. To overcome this difficulty we introduced a system of linking each fragment or column with specific metadata by virtue of a `corresp` attribute (Fig. 6). This allowed us to include links to the more than 6,000 images of Herculaneum papyri or their apographs that are available on the internet.

Fig. 6: XML of TM 60411.

By virtue of the `corresp` attribute the data about an image is then displayed together with the text of the fragment it depicts (Fig. 7). This system is followed throughout in order to link a fragment with specific image files, whereas *Images* in the header refers to an internet resource (mostly the site of a papyrus collection) where images are available. In the screenshot, one can see under the title the inventory number of the papyrus and its corresponding fragment, as well as information and links to images of various types. Furthermore, in fragment 4 there is additional markup at the beginning of the line: the lower square brackets mark text supplied on the basis of a parallel tradition, which has been drawn on in order to supplement parts lost in the original.

Fig. 7: display of TM 62499, Philodemus *On rhetoric* 2.

Digitization of the Herculaneum papyri was greatly facilitated by collaboration with Daniel Riaňo, who is using his software *AristarchusX* for grammatical analysis and author recognition in the Herculaneum papyri.[11] He transcoded all the texts previously available at the *Thesaurus Herculanensium Voluminum* and many others. His work yielded further improvements, such as automatic lemmatization and spell-checking of the text. At a first stage we had included lemmata and numbered single words in order to allow linking to his software. However, this further level of information created problems both to the human reader of the XML and to the transformation stylesheets of the Leiden+ converter. We thus decided to suppress it for the time being while exploring solutions with stand-off markup.

3 *Anagnosis*

The *Anagnosis* project may be considered the first project user of DCLP as it natively builds upon the texts encoded there while it also aims at expanding and enlarging coverage of ancient authors. Being part of the *Kallimachos* project (http://kallimachos.de), whose objective it is to create a regional centre for digital edition and quantitative analysis in Würzburg, *Anagnosis* is working on image analysis. The aim is to link illustrations and transcriptions of papyri or, more precisely and technically, to link the edited texts of Greek literary papyri from the full text database of the *Digital Corpus of Literary Papyri* with the digital images of the originals provided by the individual collections.

---

11 See RIAÑO RUFILANCHAS 2014.

By default, the link under *Images* in the metadata taken from LDAB leads to the page of the respective repository or collection that provides the images. In total, there are some 12,000 digital illustrations available covering approximately 10,000 Greek literary papyri (TM numbers). However, the distribution is uneven: while no images of many literary papyri are available, there can be hundreds of a single Herculaneum papyrus, reproducing different fragments and columns, which were taken at different times and with different methods. Thus 152 Herculaneum papyri are covered by 6,000 illustrations, and more than 3,000 of them are engravings from the 18th and 19th centuries. It was partly for the needs of and thanks to the data provided by *Anagnosis* that the new display of links to individual images above single fragments was introduced. This makes it possible to automatically create a parallel display of image and text (Fig. 8).
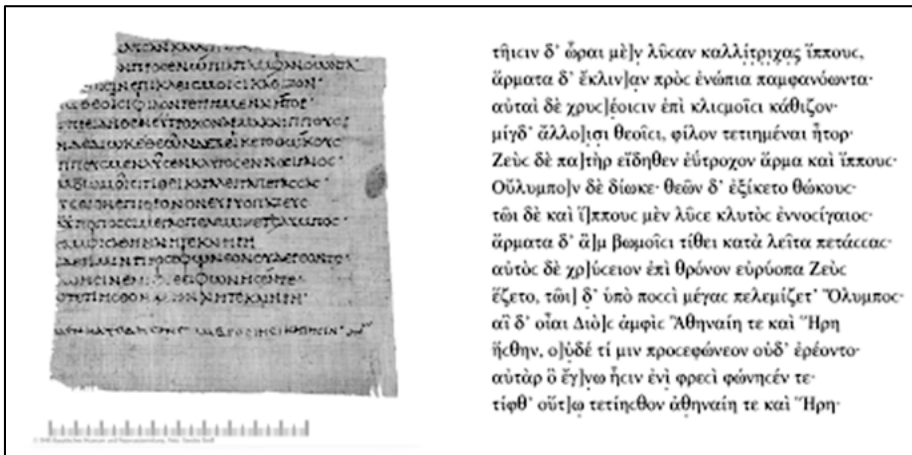
Fig. 8: TM 60411: Homer, *Iliad* VIII 433–47. Image © BerlPap. Berliner Papyrusdatenbank.

The output is created entirely from separate digital sources and is loaded *ad hoc* for further processing. The image is retrieved from the server of *BerlPap*, a project of the Berlin Papyrus Collection, whereas the transcript comes from DCLP. Thus *Anagnosis* links the two resources by virtue of the TM number and the fragment number, if applicable. As explained above, these links are stored in the corresp-attribute in DCLP.

Some might find already this parallel display useful, but in the perspective of *Anagnosis* this represents just the starting point of image analysis, with the ultimate goal being to link each letter of the transcript to the corresponding zone in the image. This will allow us to extract and display a complete alphabet of the letters present in the papyrus, a tool traditionally used for deciding on the reading of damaged areas. Such a complete set of images for each letter will further enable automated examination of

the uniformity of shapes and of the scribe's consistency. In addition, there is the possibility of automated graphic reconstruction to evaluate the spacing of proposed supplements.

The development of the underlying software was carried out at the German Research Center for Artificial Intelligence by Saqib Bukhari and will be published in specialized journals related to image analysis.[12] The software will be freely accessible for use and available for download. *Anagnosis* will only produce snippets of single letters and coordinates in a stand-off TEI format. Thus by working with the software the user is not requested to give away images of original papyri, but he may contribute by enlarging the pool of letter snippets that may be used for further research.

Thus the aim of *Anagnosis* is not classical OCR on illustrations of manuscripts, but exactly the opposite: starting from a line-exact copy, as it is present in the *Digital Corpus of Literary Papyri*, the corresponding letter zones within the illustration are referenced to the known letters of the copy. The main reason for this is, of course, the many technical difficulties that stand in the way of automated text recognition of handwriting and for which there is still no satisfactory solution. We therefore use the existing material to go a simpler way first. At the same time, *Anagnosis* works with the same components that are essential for character recognition: the original image, the mapping vectors and the transcript. The only difference is in the direction of the assignment. *Anagnosis* thus creates a large annotated corpus of correct OCR results from papyri, even though these results came about through detours. This corpus may then be used in future projects as training data for machine learning.

4 Bibliography

GIGANTE, M. (1979), *Catalogo dei Papiri Ercolanesi*, Napoli.

JANKO, R. (2003), ed., *Philodemus. On Poems. Book one*, Oxford.

OBBINK, D.D. (1996), ed., *Philodemus. On Piety*, Oxford.

PORDOMINGO, F. (2013), *Antologías de época helenística en papiro*, Firenze.

REGGIANI, N. (2017), *Digital Papyrology I: Methods, Tools and Trends*, Berlin – Boston.

RIAÑO RUFILANCHAS, D. (2014), *The 'Grammatically Annotated Philodemus' Project*, CErc 44,155–65.

VIERECK, P. (1925), *Drei Ostraka des Berliner Museums*, in *Raccolta di scritti in onore di Giacomo Lumbroso (1844–1925)*, Milano, 253–9.

---

12 Cf. Martin Jenckel, Syed Saqib Bukhari, Andreas Dengel: "anyOCR: A Sequence Learning Based OCR System for Unlabeled Historical Documents," 23nd International Conference on Pattern Recognition (Mexico 2016).

Lajos Berkes

Perspectives and Challenges in Editing Documentary Papyri Online

A Report on Born-Digital Editions through *Papyri.info*

1 Text editions at *Papyri.info*

This contribution reports on the digital publication of Greek documentary papyri through the platform *Papyri.info*:

> Papyri.info has two primary components. The Papyrological Navigator (PN) supports searching, browsing, and aggregation of ancient papyrological documents and related materials; the Papyrological Editor (PE) enables multi-author, version controlled, peer reviewed scholarly curation of papyrological texts, translations, commentary, scholarly metadata, institutional catalog records, bibliography, and images.[1]

The PE allows registered users to contribute content to the database. The submissions are then vetted by members of the editorial board. These contributions vary from correcting typos or entering translations to proposing new readings in already published texts. After this feature was introduced, the submission of corrections led quickly to the realization that in some cases extensive corrections may well justify reeditions of texts. This was a further impetus to experiment with accommodating born-digital editions of documentary papyri at *Papyri.info*.[2]

Six papyri have been published so far in *Papyri.info* and some more will appear soon. The first edition dates back to 2013, when Nikos Litinas transcribed the back of P.Corn. 34 via *Papyri.info*. Subsequently and after much consideration, the editorial board of *Papyri.info* decided to present his readings not as corrections, but as a proper online edition.[3] In 2014, the idea of exploring this issue in the framework of a seminar at the University of Heidelberg arose. In a class on digital papyrology offered in the summer semester 2014 by Rodney Ast, James M.S. Cowey, and myself several unpublished papyri were discussed and prepared for a digital edition. These were Gothenburg papyri that were only described in H. Frisk's 1929 publication (P.Got.). The editions were peer-reviewed by at least two specialists in the field and appeared exclusively in *Papyri.info*. The three texts were fragmentary letters from the VI–VII c.

---

1 http://papyri.info. For a detailed discussion see REGGIANI 2017, 222 ff.

2 Born-digital editions were already included in the original concept of *Papyri.info*.

3 DDbDP 2013 1: "Account of Barley and Wheat": http://papyri.info/ddbdp/ddbdp;2013;1.

AD.[4] The publication of these documents was announced in the *Bulletin of Online Emendations to Papyri* 5.1 (January 7, 2016).[5] More recently, two *ostraka* were republished in *Papyri.info*.[6] The *descriptum* O.Did. 37 was emended by Hélène Cuvigny to an extent that justified a new edition, although no introduction or line-by-line commentary was added.[7] O.Berenike II 237 was republished with introduction and commentary by Roger Bagnall and Rodney Ast.[8]

Work on these editions was a challenging task and not only for the usual scholarly reasons. Although using databases has very much become an indispensable tool for papyrologists, producing an edition which appears exclusively online but conforms to the well-established form of papyrological editions turned out to be quite difficult. In what follows, I will outline the main problems we faced during this process and present the preliminary solutions that we have been able to come up with. The form of online editions of documentary papyri is far from final: which conventions and solutions will be adopted depends very much on the reactions of the community. In this spirit, this contributions aims also at encouraging further discussion of online editions in the papyrological community.

2 An example: DDbDP 2015 1

Let us have a look at one of the editions, DDbDP 2015 1[9] (= P.Got. 54 *descr*.). The layout of this edition is essentially the same as in the case of already published texts in DDbDP. The only major difference is the presence of an extensive introduction and commentary.[10] The edition begins with a summary of the HGV, TM and APIS metadata.[11] Next follows the introduction, the image and the translation in a convenient layout.[12] The edition ends with the text and apparatus and notes to individual lines (see Figg. 1–2–3).

---

4 DDbDP 2015 1: R. AST –L. BERKES, "A Letter from Athanasios the Scholastikos Mentioning Constantinople" (http://papyri.info/ddbdp/ddbdp;2015;1); DDbDP 2015 2: A BERNINI, "A Requisition of Workmen and Donkeys" (http://papyri.info/ddbdp/ddbdp;2015;2); DDbDP 2015 3: L. BERKES, "A Wife in Prison: A Letter from 7th Century Fayum" (http://papyri.info/ddbdp/ddbdp;2015;3).

5 AST – BERKES – COWEY 2016. Cf. REGGIANI 2017, 241.

6 These were announced (together with DDbDP 2013 1) in AST – BERKES – COWEY 2017.

7 DDbDP 2016 1: http://papyri.info/ddbdp/ddbdp;2016;1.

8 DDbDP 2016 2: http://papyri.info/ddbdp/ddbdp;2016;2.

9 For discussion of the criteria behind this referencing system see below.

10 It is possible in the DDbDP to add introduction and line-by-line comments to any published papyrus as well: cf. REGGIANI 2017, 241.

11 The content of APIS has been integrated into *Papyri.info*.

12 The layout of the edition may depend on browser settings.

Fig. 1: The presentation of metadata.

Fig. 2: The layout of the edition: introduction, image, and translation.

r

† τὰ ἀμφισβητηθέντα ἕνεκεν τ[-ca.?-]
τῇ εἰκάδι τοῦ Χοίακ μη[ν]ὸς εἰς τὴ[ν -ca.?-]
ὑπεροχὴ(") τὰς δ[ι]αλίψεις(") καὶ το[-ca.?-]
κεινηθέντας(") ἐν τῷ κεφαλαίῳ τ[-ca.?-]
5   σημᾶναι, vac. ? πλὴν θεοῦ δεδωκότο[ς -ca.?-]
τῆς μεθοδίας(") κατὰ Ὑπατίου τοῦ ἐν[-ca.?-]
ἐκπέμψαι κατὰ τὴν βασιλίδα τῶν π[όλεων -ca.?-]
-- -- -- -- -- -- -- -- -- --

v

[-ca.?-] † Ἀθανάσιος σὺν θ(εῷ) σχολ(αστικὸς) ὑμέτερ(ος)

Apparatus

^ r.3. or ὑπεροχῇ
^ r.3. l. δ[ι]αλήψεις
^ r.4. l. κινηθέντας
^ r.6. l. μεθοδείας

Notes

2·

Cf. D. Hagedorn, APF 59 (2013) 123–137, esp. 127-136.

3·

ὑπεροχή (or ὑπεροχῇ) was probably proceeded by ἡ ὑμετέρα in the previous line, either in the nominative or dative case, see, e.g., CPR 14 9.10 (Arsinoite, July - Nov. 607); P.Münch. 3.1 79.5 (prov. unknown, 6th/7th c.); P. Oxy. 63 4397.192, 219 (March 17, 545). The word refers to high-ranking persons, usually duces, cf. J. Gascou, [Review of] P.Mich. XIII, CdÉ 52 (1977) 360–368, at p. 363, and G. Agosti, AnalPap 10-11 (1998-1999) 7–15, at p. 13. Instead of διαλήψεις, διαλείψεις ("intervals") might have been intended. We prefer the former, however, because it is more consistent with the prevailing theme of legal procedures (see intro above).

Fig. 3: The layout of the edition: transcription, apparatus, and commentary.

This structure preserves the traditional, well-established format of papyrological editions, but offers some distinct practical advantages. Scholarly work – even in the humanities – has been heavily relying on digital resources over the past decades: research is basically unthinkable without the extensive use of databases, journals and books available online. This is especially true for papyrology, since the field offers a wide range of excellent and indispensable databases for its practitioners. Furthermore, engaging with the discipline often requires dealing with scattered information: where to find an image of the papyrus? Which textual corrections or interpretative suggestions have been made to the ancient texts one is working on? Finding this information is relatively easy in the very well organized field of Papyrology, but after finding the references, scholars often end up searching other databases for the referenced images, texts, articles, and books online.[13]

---

13 For discussion of the methodological advantages and the development of papyrological digital resources, cf. Reggiani 2017, *passim.*

The digital editions offered in the DDbDP clearly faciliate this process. The (downloadable and zoomable) image appears next to the papyrus, which renders checking the readings very easy.[14] Metadata and links to the relevant entries of other important databases in the field can be found in the introduction, while an effort has been made to give direct links to all papyrological texts in the DDbDP and the referenced articles,[15] when an online version is available – which is increasingly the case. The reader gets all the information, which he or she would otherwise collect from different sources, in one package.

3 Technical challenges

I have tried to demonstrate the advantages of digital editions by using the example of DDbDP 2015 1. However, it has to be said that producing these editions was a difficult undertaking often hampered by technical challenges. Sometimes the display does not show the result one would expect even though the XML markup is correct. There were numerous problems in laying out the texts, which often required *ad hoc* solutions. An example are raised letters which are often used to designate second editions. However, these cannot be used in the commentary or the introduction: they have to be represented in other ways. A more serious problem is the case of PDF files: one of the most important features, which the scholarly community would expect, is that online editions can be converted into downloadable PDF files, i.e. print versions. This way digital editions could be accessible in a more traditional, tangible form. It seems, however, that this is technically a much more complex issue than one would expect. There is unfortunately no easy way at present to transform the digital editions made at DDbDP into PDF files.

There are also limitations on encoding those features in Leiden+/XML that were already difficult to handle during the digitization of print publications. It is very difficult, for example, to reproduce the layout of the papyri at present. This is not that visible in the case of papyri published in the DDbDP so far, but there would be certain cases (e.g. accounts with specific layout) that would be very difficult to reproduce. It is also impossible at the moment to represent abbreviations in the apparatus, as in printed editions. The abbreviations of the address in DDbDP 2015 1 ( ] † Ἀθανάσιος σὺν θ(εῷ) σχολ(αστικὸς) ὑμέτερ(ος)) would have been represented in the apparatus of a printed edition in the following manner: σ^{θ}υνσχ^{λ}ουμετερ*ρ* *pap*. Whether one includes

---

14 The images are of course hosted by the Gothenburg University Library. The online publication of images not available on the instituional websites of their respective holders and copyright owners may be an additional challenge in other cases.

15 If the whole article was not available online, the relevant entry in the online *Bibliographie Papyrologique* via http://papyri.info/bibliosearch was linked.

this in the apparatus is very much up to editorial traditions and preferences; one may argue, in fact, that the presence of a high quality image of the papyrus renders detailed approximations of abbreviations superfluous. However, a digital edition *could* offer more in this case: it would be certainly possible to link the abbreviations in the text directly to the image. [16] Even individual letters of the transcription could be matched with the image and thus online editions could become an excellent tool for self-study.[17]

These technical issues have imposed some limitations on the editions published in the DDbDP, but there is no doubt that all these problems *could* be solved. If such digital editions are be accepted by papyrologists, it will be only a question of time and money to create an editorial platform at *Papyri.info* that can serve all the needs of traditional editions. The question is rather where the priorities of papyrologists lie, or in other words: is this worth the effort in a field with such limited resources? Another question may be asked at this point as well: should we expect online editions to conform to the norms of traditional printed editions or should we accept them as a slightly different form of publication?

4 Referencing digital editions

One of the most difficult problems we faced after creating the editions was their naming and referencing. There are two major issues here: 1. How should these editions be integrated in the standard reference system of papyrology, *Checklist of Editions of Greek, Latin, Demotic, and Coptic Papyri, Ostraca, and Tablets* (hereafter *Checklist*)?[18] 2. Since these texts are part of *Papyri.info*, their text, introduction and commentary can be emended and updated by users and these improvements may change the 'original' edition. This would imply that these editions have no stable form, but are on some level fluid.[19] This 'fluidity' obviously creates difficulties in referencing: how can this problem be dealt with?

Both issues have created significant discussion among editors of the DDbDP and the solutions are preliminary. In what follows, I will outline the main lines of thinking

---

16 This method is being applied on literary papyri in the *Anagnosis* project: http://www.kallimachos. de/kallimachos/index.php/Anagnosis (see the chapter by R. Ast and H. Essler in the present volume).
17 For the usefulness of online paleographies cf. *PapPal* (http://www.pappal.info/). For an online papyrological school see the online *Arabic Papyrology School* (http://www.apd.gwi.uni-muenchen. de:8080/aps/home/), which uses the method of matching letter forms with the transcription in order to introduce students to the paleography of Arabic papyri.
18 http://papyri.info/docs/checklist. See REGGIANI 2017, 23 ff. and 29 ff. on the *Checklist* and bibliographical standards in Papyrology.
19 Cf. REGGIANI 2017, 241.

behind the solutions which have been implemented so far. Since editors themselves disagree on some of these questions, it has been decided to wait for the reactions of the papyrological community and continue developing the form of the editions based on its reactions.

The papyri which have been published in the DDbDP so far were all described before and had a corresponding reference in the *Checklist*. This made things easier in a way, since we could have opted for slightly modifying the already existing references of these papyri, but at the same time complicated issues even more, since we did not want to introduce anomalies in the *Checklist*. Furthermore, we also had to consider a reference system for papyri that did not have a *Checklist*-identifier yet, so that it could be used for further publications.

At first sight, these publications represent the same case as papyri published in journals. They could have been referenced in a similar fashion and then included in the *Sammelbuch griechischer Urkunden aus Ägypten*, as usual. However, an important argument came up quickly: why should we double the effort, if the texts printed in the *Sammelbuch* are reentered into the DDbDP again? This is especially true considering that it is inevitable in the long run that the *Sammelbuch* will be published digitally (only?). It was also important to indicate the date of publications, since this way these editions become comparable to journal publications. Taking all these considerations into account it was decided to use the identifier *DDbDP year number*, e.g. DDbDP 2015 1: this identifier represents a collection of digital-only publications.[20] The numbering follows the sequence in which the submitted publications were included in the DDbDP. This creates a clear and transparent system that enables straightforward referencing of editions born at *Papyri.info*. These editions are regularly announced through the *Bulletin of Online Emendations to Papyri* (BOEP)[21] in order to make their existence known to the papyrological community.

The other issue is what I have called the 'fluidity' of these texts. *Papyri.info* enables editing the text, introduction, and commentary – essentially any part of these digital editions. This leads to obvious problems in referencing them. For instance, let us say someone refers in an article published in 2018 to the introduction of DDbDP 2015 1. However, in 2019 a user of the DDbDP proposes a change to the introduction of DDbDP 2015 1 that affects exactly the part of the text which was referred to in 2018. His or her suggestion is reviewed and accepted by the editorial board of the DDbDP and subsequently replaces the text referred to in the 2018 article. If someone checks the 2018 article in 2020 and wants to look up the reference, he or she will not find it. The information that the introduction has changed would be available in the editorial

---

20 It has not been decided yet whether these publications will be included in the *Sammelbuch* or not.
21 Edited by R. Ast, L. Berkes, and J. M.S. Cowey: http://www.uni-heidelberg.de/fakultaeten/philosophie/zaw/papy/projekt/bulletin.html

history of DDbDP 2015 1, but this is not an obvious or user-friendly place to look for this information.

This is a problem at the present, even if it is only a theoretical one, since no changes have been made to these editions so far. The editorial board of the DDbDP has debated how to deal with such scenarios, but there are basically two solutions. The more traditional way is to track the different versions of these editions. This means that in theory something like DDbDP 2015 1 (2), 1 (3) etc. would come into existence each time the edition would change. In an ideal world the user would be able to switch between these versions with the differences being highlighted. However, this is impossible to do presently.

Another way of thinking would be to accept these digital editions as a new form of scholarly publication that is not as stable as the traditional ones. This would imply that scholars would need to get used to the idea that the texts they find online can change anytime and that they need to check them each time before quoting them and always refer to them with a date of access. This approach has certainly some appeal, as it creates a faster, more direct way to do scholarly work, but there are also *caveat*-s. This method may lead to chaotic references and a certain lack of transparency. However, we also have to accept that at some level this is already becoming the reality of scholarly work. Discussing drafts publicly on an online platform (e.g. at https://www.academia.edu) has become increasingly common even in the humanities (this practice is much more widespread in the sciences). We all refer in papyrology to unpublished documents, drafts, personal communications of colleagues: in the past, even communications "per litteram" were normally cited and accepted as scholarly references. These are in a way also messy references, since they cannot be easily checked or verified, especially since some of this information never becomes publicly available. If we want to fully exploit the possibilities of digital publications, we may need to accept this 'fluidity'.

Finally, I would like to mention a further minor problem in referencing the text: there is no traditional page numbering. In my view, however, this is not a real problem. First of all it is very common in Papyrology to refer only to introductions or commentaries of editions without mentioning the page number. But what is more important: references can be very quickly found in an online environment through the search function of the browser. Even though some references to online editions may seem vague at first, it is very easy to deal with them practically.

```
Editorial History; All History; (detailed)
2016-03-10T10:16:41-05:00 [james.cowey]: Finalized - Ready.
2016-03-10T09:57:36-05:00 [james.cowey]: Vote - Accept-Straight-to-Finalization -
Fine
2016-03-10T09:55:57-05:00 [james.cowey]: Submit - For renaming.
2015-12-16T09:23:00-05:00 [simoeis]: Finalized - Ready
2015-12-16T09:19:11-05:00 [simoeis]: Vote - Accept-Straight-to-Finalization - Fine
2015-12-15T10:55:21-05:00 [simoeis]: Submit - listBibl has been removed
2015-06-11T08:55:16.459-04:00 [james.cowey]: Finalized - Ready. This edition has
undergone the peer review process having been vetted by two senior editors. Their
suggestions and corrections have been taken into account.
2015-06-11T08:55:16.451-04:00 [james.cowey]: Vote - Accept-Straight-to-
Finalization - Fine.
2015-06-11T08:46:09.179-04:00 [james.cowey]: Resubmitting an edition prepared by
R. Ast and L. Berkes, because problems with finalization had arisen
2015-05-28T08:30:40.46-04:00 [berkes.lajos]: We are submitting a full edition of
this papyrus, including revised text, HGV data, commentary, and translation. We have
taken into account suggestions made by the readers.
```

Fig. 4: The editorial history of DDbDP 2015 1.

5 Perspectives of digital editions

I have tried to demonstrate the main advantages and the accompanying problems of digital editions at *Papyri.info*. As I have emphasized: it is still an open question whether this format will be accepted and developed. If it is accepted, it would certainly represent a new, more fluid form of textual editions in our field. If we try to look at the bigger picture, this model offers some distinct advantages beside the practical ones I have mentioned so far. Publication through this platform is open-access, peer-reviewed, and fast. Once an edition has been properly vetted, it can be published without further delay. This platform may also offer the possibility to quickly describe or publish smaller fragments that could swiftly enter DDbDP this way and thus become searchable. This of course should not imply that *Papyri.info* limits itself to the publication of small pieces that would otherwise be ignored.

There are also some problems with this model. One issue is the appreciation of online, open-access publications in our field and the humanities more generally. Even though most scholars agree that such publications are desirable and represent the future, they are still not really valued. If someone were to decide to publish a papyrus in a well-established journal or at DDbDP, it is pretty clear that the former possibility is much better for one's CV, even if submissions to the DDbDP go through a peer-review process. An additional problem in this respect is that while in an article

one can bundle several papyri together, this is not possible at present at *Papyri.info*.[22] Only time can tell how fast the attitude towards digital publications will change in our field. However, it is pretty likely that this will happen, as has been the case in the sciences for quite some time.

Another problem is the issue of limited resources in our field. *Papyri.info* is indispensable at the moment, but is struggling to keep up with new publications, BL entries and other corrections, etc.[23] So the question is: what are our priorities? The focus of the editorial board of *Papyri.info* has been very pragmatic: making as many new texts as possible available. This has led to certain compromises. For instance, alternative readings are often not entered during the entry stage and many (BL and other) corrections are still missing. We believe that it is better to have more material online (even with some infelicities) than to stick to a more limited, but also more flawless corpus.

Focusing more on publishing papyri on *Papyri.info* would certainly also require an effort from the papyrological community: scholars would need to vet the incoming submissions and be open to treating editions at *Papyri.info* the same way as they would treat printed publications. The online publication does not have to stop with documentary texts; the existence of the *Digital Corpus of Literary Papyri* (DCLP) at http://www.litpap.info could also open the door to editing literary papyri using essentially the same platform.[24] Languages do not have to be a limit either: at present it is possible to publish Latin, Greek, and Coptic papyri at *Papyri.info* and the language of publication is not restricted to English.[25] The potential is certainly there, but priorities need to be set.

I believe that the only way to find out whether *Papyri.info* could work as a platform for editing papyri is to give it more practice. We need to see whether this kind of edition and reference system works for Papyrology or not. It may turn out at the end that some (inevitable) chaos in referencing these texts is outweighed by the advantages that this kind of direct scholarly work provides; on the other hand, it is also possible that papyrologists decide to stick with more traditional forms of publication or prefer other digital options. However, to assess this we would need more online publications. At moment, anybody can submit a papyrus for publication at *Papyri.info*. It is hoped that this article will encourage scholars to explore the possibilities of editing papyri on this platform or at least to open a dialogue about its advantages and disadvantages.

---

22 On the other hand, producing a digital-only *volume* is certainly possible.

23 However, as R. Ast pointed out to me, the situation was much worse in the late 1990s and early 2000s: the difference is that users' expectations are much higher now.

24 Cf. REGGIANI 2017, 250 ff. and the chapter by R. Ast and H. Essler in the present volume.

25 In fact, the editorial histories of certain submissions at *Papyri.info* often contain discussion in French, German, Italian or even other languages.

6 Bibliography

Ast., R. – Berkes, L. – Cowey, J.M.S. (2016), eds., *Bulletin of Online Emendations to Papyri* 5.1 (January 7, 2016), Heidelberg, URL: http://www.uni-heidelberg.de/md/zaw/papy/forschung/ bullemendpap_5-1.pdf.

Ast., R. – Berkes, L. – Cowey, J.M.S. (2017), eds., *Bulletin of Online Emendations to Papyri* 6.1 (March 22, 2017), Heidelberg, URL: http://www.uni-heidelberg.de/md/zaw/papy/forschung/ bullemendpap_6.1.pdf

Reggiani, N. (2017), *Digital Papyrology I: Methods, Tools and Trends*, Berlin – New York.

Massimo Magnani

The Other Side of the River

Digital Editions of Ancient Greek Texts Involving Papyrus Witnesses

1 Introduction

More than forty years ago, on the footsteps of Dom Froger (1968), WEST 1973, 70–2 asked himself for which editorial task the use of the computer could have helped. After having discarded the automatic collation,[1] West imagined that a computer, provided with the transcriptions of the manuscripts "purged of coincidental errors", could have drawn up "a clumsy and unselective critical apparatus". Then, if contaminations have been out of question, this computer could have worked out "an 'unoriented' stemma" by comparing the variants, but it could not have been able to choose the correct orientation of the stemma, an operation possible only "by evaluating the quality of the variants".[2] Thereafter, it has been assumed that the computer might be useful even for an heavily contaminated paradosis: the late Bryan Hainsworth, presenting in short the manuscript tradition of the *Odyssey*, imagined that a computer could establish the degree of contamination of every single family of the *Odyssey*'s paradosis, but in his opinion the result would not be commensurate with the amount of work.[3] After years, what is the situation, with reference to the ancient Greek literature and texts? If we apply to the term 'edition' the usual scholarly meaning, i.e. 'edition of a text based on the method(s) of textual criticism', and we expect that the expression 'digital edition' can refer, among the variety of the editorial products, not only to the more or less refined digitization of old or new 'traditional' editions of ancient Greek texts,[4] but also to the 'edition of a text based on a new digital method of textual criticism', we are destined to disappointment, at least for the time being, and not surprisingly.

---

1 "Machines have not yet been devised which can cope with variations inherent in handwriting", p. 71. *Transkribus* promises to find a solution to the problem of the automatic transcription (https://transkribus.eu/Transkribus). For the automatic collation of transcriptions, see *CollateX* (https://collatex.net/about), the successors of the well-known Peter Robinson's *Collate*, the equally well-known Wilhelm Ott's *TUSTEP* (http://www.tustep.uni-tuebingen.de/tustep_eng.html), and finally *Juxta* (http://www.juxtasoftware.org). West's 'traditional' critical edition of the *Odyssey* was recently published posthumously (WEST 2017).
2 WEST 1973, 72, and see the example on p. 71.
3 HAINSWORTH 1982, xxxiv.
4 Texts transmitted by multiple witnesses and editions produced *via* the traditional, post-Lachmannian textual criticism.

2 'Digital editions' of classical texts: an overview

The *Catalogue of Digital Editions*, the most complete "attempt to survey and identify best practice in the field of digital scholarly editing",[5] gathers 256 'digital editions', among which 12 edition of ancient Greek texts (30.10.2017).[6] In fact, none of them is a new edition based on textual criticism applied to a multi-manuscript tradition: 11 editions are catalogued as 'digital scholarly editions';[7] six of them are editions of a text transmitted by unique witness (4 are editions of epigraphic collections,[8] two involve the text transmitted either by a single[9] or by a peculiar and very valuable ancient manuscript[10]), two are the digitization of the standard, 'traditional' editions of the *Old* and *New Testament* in Greek language (ID 163 = *LXX Septuaginta* – http://septuaginta.net; 73 = *Digital Nestle-Aland* – http://nestlealand.uni-muenster.de), one is the edition that gathers the electronic editions of the *Gospel according to St. John* in Greek, Latin, Syriac and Coptic (ID 58 = http://www.iohannes.com). Finally, I will not include the *Digital Athenaeus* among the digital scholarly editions, since it is the digitization of

---

5 A project of P. Andorfer and Ks. Zaytseva of the Austrian Centre for Digital Humanities (ACDH, Vienna) and G. Franzini of the UCL Centre for Digital Humanities (UCLDH, London); for the data, see FRANZINI 2012–; for the web site, see FRANZINI – ANDORFER – ZAYTSEVA 2016–.

6 See also the *Catalogue of Scholarly Digital Editions*, compiled by P. Sahle (http://www.digitale-edition.de). Neither the *DIG-ED-CAT* nor Sahle's *Catalogue* register the *Homer Multitext Project* (http://www.homermultitext.org). On the subject of the digital scholarly editions, see in general PIERAZZO 2015.

7 *DIG-ED-CAT* ID 10, 24, 57, 58, 73, 86, 92, 93, 100, 163, 244. The ID 17, the edition of the *Euripides scholia* managed by D.J. Mastronarde (http://euripidesscholia.org/EurSchHome.html), started in 2007 and currently updated in November 2017, is not included in the catalogue of the digital scholarly editions not because this edition is not digital, but because it is credited not to be a scholarly edition, term by which the *DIG-ED-CAT* project managers "mean editions with a strong critical component" (https://dig-ed-cat.acdh.oeaw.ac.at/faq). I am not able to understand why the *Euripides scholia* are not a 'scholarly' edition, but an edition like *Sappho's poems* (http://inamidst.com/stuff/sappho), "an attempt to collect Sappho's entire work together in one page — with Greek originals, succinct [?] translations, and commentary [in fact, absent]", it is. Obviously, these catalogues cannot provide any guarantee of completeness.

8 ID 24 (IOSPE = *Inscriptions of the Northern Black Sea* – http://iospe.kcl.ac.uk/corpus/index.html), 86 (IGCR = *Inscriptiones Graecae in Croatia repertae* – http://www.ffzg.unizg.hr/klafil/dokuwiki/doku.php/z:epidoc-hrvatska), 92 (InsAph = *Inscriptions of Aphrodisias Project* – http://insaph.kcl.ac.uk/index.html), 93 (IRCyr = *Inscriptions of Roman Cyrenaica* – https://ircyr.kcl.ac.uk). For the s.c. "special types of edition", i.e. the edition of papyri and inscriptions, see WEST 1973, 94–5.

9 ID 57 (*Derveni Papyrus* – https://chs.harvard.edu/CHS/article/display/5418).

10 ID 10 (*Codex Sinaiticus* – http://codexsinaiticus.org/en). The *DIG-ED-CAT* does not include a digital scholarly edition very similar to the *Codex Sinaiticus*, that is *Palamedes* (*PALimpsestorum Aetatis Mediae EDitiones Et Studia* – http://www.palamedes.uni-goettingen.de), the edition of the palimpsest mss. *Hierosol. Sancti Sepulcri* 36 and *Par. Gr.* 1330.

Kaibel's text (with tools).[11] The same situation has been recently underlined by Ermanno Malaspina with reference not only to the ancient Greek domain but also to the philological studies in general.[12] An exception, once finalized, should be his project of digital edition of the Ciceronian *Lucullus* (in collaboration with *MeDiHum*, Turin University, e IAS, Durham University): a 'Lachmannian' critical edition of a classical text transmitted by 74 mss. arranged in a closed recension, a sort of 'crash test' of the DH resources, and most of all of the TEI encoding.[13] The final result will be a web page with critical text and apparatus, with the possibility to open windows with the text of individual manuscripts or editions in correspondence with their variant readings of ecdotic significance. To mark a difference from the mere digitization of a 'traditional' critical edition, the apparatus of this *Lucullus* online aims to be much richer in data without yielding to the principles of genetic philology, but offering all that is relevant for the tradition of the text. In 2015 Malaspina performed a transcript from Word to XML through *Oxygen* of the readings of the 74 manuscripts and of some printed editions limited to 4 text-paragraphs. By following the TEI guidelines (ch. 12), each variant reading was tagged (`<rdg>`) referring through the `@wit` attribute to the list of the mss., and through one of the four `@type` attributes to its ecdotic classification (in order of increasing relevance: orthographical variants, polygenetic variants, variants significant for the constitution of the text, variants significant for establishing the

---

11 ID 244 (http://digitalathenaeus.org). The *Digital Athenaeus* is a "work […] focused on annotating quotations and text reuses in the *Deipnosophists* in order to […] provide an inventory of authors and works cited by Athenaeus, and to implement a data model for identifying, analyzing, and citing uniquely instances of text reuse in the *Deipnosophists*", where the Greek text is the digitization of the Teubner edition of Kaibel 1887–90 without critical apparatus (see also Berti – Daniels – Strickland – Vincent-Dobbins 2016; Berti – Blackwell – Daniels – Strickland – Vincent-Dobbins 2016). S.D. Olson is at work, in order to produce a new, 'traditional' critical text of the *Deipnosophists* (the first volume of this edition will be published in 2018).

12 Malaspina in Malaspina – Della Calce 2017, 58–9: "Con la formula 'edizione digitale' oggi si intende praticamente di tutto: riproduzioni di epigrafi, scansioni di brogliacci, collazioni di varianti e così via. Anche se si aggiunge l'aggettivo 'critica', che nella filologia classica tradizionale indicherebbe un prodotto ben definito, non si ottiene un quadro più omogeneo e soprattutto si vede spesso gabellato per 'critico' ciò che sarebbe più onesto definire 'diplomatico', ovvero la mera trascrizione di un testimone e/o la giustapposizione di varianti senza nemmeno porsi il problema della *vera lectio*". With reference to the Romance studies, Rinoldi in Bernagou – Palumbo – Rinoldi 2016, 41–4 underlines some consequences of the increasing online diffusion of the 'virtual manuscripts', if compared to the lack of digital critical editions: even though the application of this approach should be ideally restricted to documentary texts, texts transmitted by a single ms., and particularly venerable mss., the alignment of reproductions and transcripts of mss. could promote the return of the bad Bédierism, that is the fetish of the *bon manuscrit* (see below). See Pierazzo 2011 on the digital scholarly edition of documentary texts.

13 Malaspina in Malaspina – Della Calce 2017, 58–62.

stemma codicum). The *JavaScript* prototype, performed by Peter Heslin (*Diogenes*' inventor), allows for the data display and the control of the tagging errors; the last stage will be the synoptic display of the text of each witness.

The lack of "'comprehensively digital' scholarly edition of a 'classical' text with a manuscript-based multi-testimonial tradition" was noticed by Monella 2012.[14] According to the responses given to his cognitive survey on *Academia.edu* and *Digital Classicist*, the main reason should be "time and money", but Monella believes that the real reason is the lack of need. Digital scholarly editions (DSE) are favourite scholarly products of codicologists, epigraphists, papyrologists, editors of documentary manuscripts and palaeographers, says Monella, because they are focused on documents, and by

> 'genetic' editors of modern and contemporary texts [...][15] and historical linguists, who may study the evolution of language and orthography through 'errors' in inscriptions, in manuscripts and in modern print materials throughout the centuries,

because their concern is the textual variance. Classical editors who are dealing with texts transmitted by a "manuscript-based multi-testimonial tradition" are dealing with 'canonized' ancient texts, where textual variance is due – or credited to be due – for the most part to the erroneous medieval paradosis and not to the author. 'Errors' are identified and collected only for establishing the *stemma codicum* and the text itself. Finally, continues Monella, for classical editors the manuscript is important especially as 'textual vehicle' and does not bear a particular relevance in itself. Therefore, e.g., why should the *Aeneid*'s editor digitize even a limited part of this manuscript tradition? The purpose should not be the expansion of the critical apparatuses with the inclusion of more data – usually, this purpose is disregarded by the 'traditional' editor because of their editorial irrelevance, but "a plural, fluid concept of text, a concept implying that each document's text is worth studying as a historically determined cultural object", and the increasing interest in "'post-classical' Latin and Greek – thus joining forces with historical linguists and romance philologists". In my opinion, it is not completely true that the 'traditional' classical editors and philologists are only interested in manuscripts as witnesses of the text: on the contrary, we see an ever-growing attention to the ancient and medieval manuscripts as essential witnesses of the cultural reception of the related texts. It is acceptable that the digital scholarly editions can provide the best framework, in order to manage and display

---

14 See also Tomasi in Italia – Tomasi 2014, 120, noticing the need of shared criteria such as verifiability of the sources, reliability of the institution promoting the edition, presence of curators, scientific layout, dates of creation and updating, absence of commercial interference (and, hopefully, absence of mistakes). In her opinion, to the absence of shared criteria, it is to be added that the DH are sometimes perceived by philologists as a mere instrument useful for speeding up procedures rather than as a new way of understanding the edition, and that the interface often hides the methodological accuracy that has governed the process of creating the edition.

15 See e.g. D'Iorio 2010; Italia in Italia – Tomasi 2014, 128–30, especially on the analytic genetic editions.

the rich data complex derived by a multi-manuscript paradosis in its entirety (see also below). Certainly, this approach can be of great cultural importance and can stimulate the discipline to reconsider its goals and methodology;[16] moreover, the IT solutions devised in response to the problems associated with digital editions constitute an advancement and it will contribute to create new skills and new professional figures, e.g. the 'computational scholars', that is "philologists skilled in both classical philology and computer science".[17]

3 Case study 1: Mastronarde's *Scholia Euripidea*

Before trying a provisional conclusion, I would like to review with some detail two digital scholarly editions. Among the aforementioned editions, the only one that is really a new edition, even though based on the 'traditional' method(s) of textual criticism, are Mastronarde's *Scholia Euripidea* (*supra*, n. 7), aiming to supersede the standard work of SCHWARTZ 1887–91.[18] As Mastronarde writes, "the goals of this project are quite traditional in a philological sense, but also experimental and forward-looking in terms of format". This view is very instructive. On the one hand, Mastronarde did not use the computer resources, in order to review the collations made by the previous editors,[19] to "clarify the extent, nature, and possible stemmatic relationships of the scholia in some of the so-called recentiores", or to put a better order to the scholiastic corpora of the Palaeologan era (Maximus Planudes, Manuel Moschopoulos, Thomas Magister, Demetrius Triclinius). On the other hand, his choice for an open-access digital format responds to specified intellectual and educational needs, apart other very sensible economic, professional and scientific reasons: "a digital format is variable [...], updatable, [...], allows for sharing of interim stages of the

---

16 See MONELLA 2012, n. 35, with bibliography, reconsidering "the historical and anthropological/ethnological foundations of the discipline".

17 See e.g. BOSCHETTI 2009, 5.

18 In the update of November 2017 Mastronarde announces the online publication of the edition of the *scholia* on *Orestes* 1–500 and, in the meantime, the forthcoming open-access publication of the *Preliminary Studies on the Scholia to Euripides*. The page about the 136 *sigla* used for Euripidean manuscripts is new, with the possibility to download them as Excel spreadsheet (EurSiglaTable.html), together with an updated 'Manuscripts page' with additions.

19 The description of the manuscripts is the only part that has been substantially updated (in 2016: https://euripidesscholia.org/EurSchMSS.html; in 2017: https://euripidesscholia.org/EurSchMSS new. html). Also, Mastronarde has been able to improve Schwartz's collations of **M**, **B**, and **V**; a greater progress has been made adding some lemmas and glosses of **C**, all the *scholia* in **H** (the Jerusalem palimpsest, for which see *supra*, n. 10), ms. unknown to Schwartz, and those in **O** (Schwartz wrongly dated the ms. to 15th century, but it is now credited to be written ca. in 1175). Mastronarde does not include for the moment the ancient manuscripts with marginal and interlinear notations, for which see MCNAMEE 2007, 253–7.

work, [...] is expandable, [...] searchable in a way that a printed volume is not". The project is so far limited to the first 500 lines of Euripides' *Orestes* and to 20 "witnesses of various kinds" of it; the play was selected by Mastronarde first because,

> as a triad play, it provides the maximum degree of variety and complexity in the annotation tradition and therefore forces one to confront most or all of the issues that may arise as the work proceeds",

then for the availability of images. The problems of conversion of a 'traditional' critical edition to XML/TEI format have been so relevant, that only *Orestes* 1–25 and 401–25 have been published, with the Triclinian metrical *scholia* to the *parodos* and the prefatory material.[20]

Mastronarde created four levels based on the TEI division-type element:[21] the `div1` element, the largest one, includes all the material related to the correspondent play, including one or two `div2` elements, containing the introductory texts and the *scholia*. `div3` always has three required attributes and occasionally an optional fourth one; the first two give a complete and expandable classification of the *scholia* ("`@type = vet, rec, mosch, thom, tri, plan, vetMosch, vetThom, vetMoschThom, recMosch, recThom, recMoschThom, moschThom; @subtype = exeg, gloss, paraphr, gram, artGloss` ["a gloss that consists only of the article agreeing with the glossed word"], `etaGloss` ["an eta placed over a Doric alpha in a lyric passage to indicate the normal form"]"). The third one required attribute is the `@xml-id` of the play. The `div4` elements are the "children of each scholion `div3`", the only one of these being mandatory is the one including the text of a single *scholion* with its lemma and its witness list (`@type` of '`schText`'). One of the main problems has been the impossibility "to key an apparatus item to a line number", problem that can not be easily overcome, considering that "anchoring each apparatus item to a single word or phrase is possible, but the markup would be far too time-consuming and in my opinion out of proportion to any possible gain". After the text of the *scholion*, "a required `seg` with `@type` of 'witnesses' contains the sigla of the manuscripts that contain the scholion", then follow seven (or less) other kinds of `div4`: `engTrans` (only for a choice of *scholia*), `lemmaPosNote` ("details about variations in the lemma, the presence of reference symbols linking the scholion to the text, and

---

20 Not always the XML method is accepted: Schmidt's *Ecdosis* (Sᴄʜᴍɪᴅᴛ 2016, 100–1), a back-to-front development model providing a user-driven framework, aims to create, publish and share digital scholarly editions without using XML, seriously affected by "the problem of 'markup variability' – the tendency by different encoders to mark up the same features in different ways". So, "instead of the linear text of XML, with embedded tags designed to apply abstract formats, *Ecdosis* uses a non-linear text and separate markup to describe textual properties, which are not arranged hierarchically, as in XML, but are allowed to overlap". For the TEI encoding limits, see Cᴜꜰᴀʟᴏ – Mᴜɢɢɪᴛᴛᴜ 2016, 89–91; in many of his contributions Schmidt underlines the structural inadequacy of the embedded generalized markup for cultural heritage texts (see e.g. Sᴄʜᴍɪᴅᴛ 2010), and it is no coincidence that XML / TEI are not always adopted for these purposes.

21 See at https://euripidesscholia.org/EurSchStructure.html.

the position of the note"), `appCrit` and `appCrit2` (the second critical apparatus, "for orthographica and other minor matters"), and `commentSim` (commentary and *similia*). Interesting is the creation of two tags systematically identifying two types of information, that in the 'traditional' editions are rare and scattered among the mss. readings: `collNotes` ("collation notes, record of difficulties in reading images, of divergences from previous reports, and reminders to check the original or a better image at some future date") and keywords, in fact a reminder "for additional description of the content in aid of searching or indexing at a later point". Both of them are not publicly displayed, but reserved to the author and collaborators for future work. Another interesting piece of information, again not always systematically recorded in the traditional critical apparatuses, is the location of the *scholia* (above the line, marginal, or intermarginal), the variation of their sequence, and the indication of the point where a *scholion* begins and the other ends. Mastronarde's choice is due to the difficulty of using superscripts in XML, therefore reserved to indicate different hands or "different versions of the same note at different locations in the same witness". That Mastronarde's edition is traditional and digital only for having chosen this format of displaying the textual contents is also clear from the treatment of the metrical *scholia*, limited to Triclinius' *scholia* on the *parodos* of *Orestes*. As Mastronarde's underlines, by assigning a different tagging to the metrical *scholia*, XML allows to display the metrical scheme and the text of Triclinius' mss. side by side with the *scholion*, while DE FAVERI 2002 had to publish them separately, at the back of the book.

Leaving aside, of course, the differences due to the progress of the studies on the manuscript tradition, a limited comparision, restricted to the *incipit* of the play (and to the *vetera* set of *scholia*), between Mastronarde's edition ("full view" mode of display) and the standard edition of Schwartz is instructive:[22] the digital edition shows five *scholia vetera* to *Or.* 1 (28 overall), all of which transmitted only by **V** (= **A** in Schwartz),[23] while Schwartz prints only three of them, with the first and the third 'tagged' with a peculiar [dia]critical sign, a *crux* indicating not editorial desperation but the recent origins of those *scholia*[24] (all of them are defined as *vetera exegetica* in Mastronarde's edition). The text of the three *scholia* in common does not differ; Mastronarde's apparatus is richer, is preceded by the English translation and has additional information (variations in the lemma, the presence of reference symbols, orthographical matters, and a brief commentary). The convenience of the digital format depends on the aforementioned statement (variability, upgradability, accessibility of interim stages of the work, expandability, searchability), but this edition did not benefit of digital tools neither for the mss. collation nor for the arrangement of the ms.

---

22 See at https://euripidesscholia.org/EurSiglaTable.html for the different *sigla* adopted by the modern editors of *scholia*.

23 Vatican, Biblioteca Apostolica Vaticana, gr. 909, ca. 1250–1280.

24 SCHWARTZ 1887, viii.

witnesses in an open or closed *stemma* designed for every single scholiastic corpus.[25] Its apparatus has been created by Mastronarde and is not the result of the information recovery based on the computer assisted collation of digital transcriptions.

4 Case study 2: the Homer Multitext Project

Returning to the initial question, Harris' warning[26] can be partially shared, but to my opinion the problem is not so much that the ordinary reader has no desire to loose him-/herself in the maze of variants that every philological operation inevitably generates, but that the concern about XML/TEI and its editorial application risks to make us lose sight of the final objective of every philological operation: that of establishing a text. On the other hand, I completely agree with him that if the application of the IT to the philology, precisely because it allows multiple choices, is transformed into the abdication of choice, this application is improper, because the job of the philologist is to choose. In a sense, the most ambitious project of digital scholarly edition is the one that has chosen to completely embrace that risk, not simply in order to avoid the choice but denying its methodological correctness. With the words of its editors, C. Dué and M. Ebbott,

> the Homer Multitext Project seeks to present the textual transmission of the *Iliad* and *Odyssey* in a historical framework. Such a framework is needed to account for the full reality of a complex medium of oral performance that underwent many changes over a long period of time. These changes, as reflected in the many texts of Homer, need to be understood in their many different historical contexts. The Homer Multitext provides ways to view these contexts both synchronically and diachronically.[27]

Therefore, the *Homer Multitext Project* offers "free access to a library of texts and images, a machine-interface to that library and its indices, and tools to allow readers to discover and engage with the Homeric tradition". Editors (and co-editors: D. Frame, L. Muellner, G. Nagy) reject the traditional approach for *Iliad* and *Odyssey* and the possibility of a critical edition of the poems establishing "an original text as it supposedly existed at the time and place of its origin".[28] Persuaded by the idea of a Homeric tradition always evolving from the pre-classical age to the Byzantine era, the

---

25 *Veteres* and witnesses before ca. 1261, *recentiores* (witnesses after ca. 1261) containing old *scholia*, witnesses for Moschopulean *scholia* on the Triad plays, witnesses for Thoman *scholia*, witnesses for Triclinian *scholia*, witnesses for Miscellaneous Palaeologan *scholia*.

26 Harris 2014, 84–5.

27 http://www.homermultitext.org/about.html; see also Dué – Ebbott 2010, 154: "unlike the standard format of printed editions, which intend to offer a reconstruction of an original text as it supposedly existed at the time and place of its origin, the Homer Multitext offers the tools for discovering, viewing, and understanding a variety of texts as they existed in a variety of time and places".

28 See Dué – Ebbott 2010, 153 and already Dué – Ebbott 2009, 5: "textual criticism as practiced is predicated on selection and 'correction' as it creates the fiction of a singular text. The digital criticism

Homer Multitext editors believe that the textual variants even of the medieval para-
dosis are the sign of the variation inherent to the system of oral poetry and that even
when the poems were finally written down, they continued to be performed orally.[29]
The consequence is that *Iliad* and *Odyssey* have never been fixed texts. Therefore, the
Homer Multitext Project is gradually editing the Homeric witnesses (papyri, medieval
manuscripts, ancient quotations), considering each manuscript and each textual var-
iation as the valuable testimony of the oral tradition. We are not told if all the manu-
scripts will be digitized: six medieval Iliadic codices of well known relevance have
been completely or partially digitized, not always successfully,[30] and only one papy-
rus.[31] That is, a work in progress. The scientific foundation of it was placed by two
'traditional' publications: "a multitext edition with essays and commentary" of *Il.* X
and an overview of the Ptolemaic papyri as witnessing the multitextuality.[32] This is
not the place where to discuss in detail the position of Dué, Ebbott and Gregory Nagy
on the Homeric tradition; nevertheless, this position is the essential reason for their
peculiar digital edition and therefore deserves a short presentation (and some con-
siderations). Bird's very compact monograph aims to be a general introduction to tex-
tual criticism applied first to classical and biblical texts (pp. 1–26), then to Homer in
general (pp. 27–60), finally, to the Ptolemaic papyri of Homer (pp. 61–100), inter-
preted as the evidence not of the "eccentricity"[33] but of the everlasting "multitextual-
ity" of the ancient Homeric tradition. Bird defines as "authentic" and "original" all

---

we are proposing for the Homer Multitext maintains the integrity of each witness to allow for contin-
ual and dynamic comparison, better reflecting the multiplicity of the textual record and of the oral
tradition".

29 The 'fluidity' of the Homeric text is in accordance with the general concept of the digital text; see
Babeu 2011, ix: "As recently as a generation ago, the 'text' in classics was most often defined as a
definitive edition, a printed artifact that was by nature static, usually edited by a single scholar, and
representing a compilation and collation of several extant variations. Today, through the power and
fluidity of digital tools, a text can mean something very different: there may be no canonical artifact,
but instead a data set of its many variations, with none accorded primacy".

30 Venezia, Biblioteca Nazionale Marciana gr. Z. 454 (coll. 0822, the famous *Venetus* A); Venezia,
Biblioteca Nazionale Marciana gr. Z. 453 (coll. 0821 = *Venetus* B); Venezia, Biblioteca Nazionale
Marciana gr. Z. 458 (coll. 0841 = Allen U4); Escorial, Real Biblioteca fonds principal y. I. 01 (Andrés
294 = Allen E3 = West E); Escorial, Real Biblioteca fonds principal Ω. I. 12 (Andrés 513 = Allen E4 =
West F); Genève, Bibliothèque de Genève fonds principal gr. 44. Only the *Veneti* A and B are com-
pletely digitized; we have also a sample of the *Genav.*, while the images of both *Scor.* are currently
unavailable.

31 The Bankes Papyrus (P.Lond.Lit. 28 = TM 60500 = M-P³ 1013 = LDAB 1623 = Allen-Sutton-West 14);
see also https://www.bl.uk/collection-items/the-bankes-homer.

32 Dué – Ebbott 2010; Bird 2010. For the following presentation of these books I am partially in-
debted with the dissertation of Martini 2013, 32–45. I take this opportunity to remember the very fruit-
ful collaboration with Isabella Andorlini in the supervision of Martini's work.

33 E.g. West 1967.

the variant readings that seem Homeric both by nature and by ancestry, but this judgment will not automatically define the other variants as "spurious" or "unoriginal": in others words, any variant reading that is not an obvious copyist's intervention must be considered authentic in principle. Synthetically, when two variants have internal and external factors on their side that prove or suggest authenticity, the Homer editor must refuse the traditional way of thinking of the philologist that would lead to make a choice between the two: "there is no need to choose one reading and reject the other".[34] This way of operating is programmatically opposed to the Lachmannian critical setting, also from the point of view of data presentation (i.e. a traditional critical edition that provides a main text at the top of the page, accompanied by a presentation of variant readings in the critical apparatus at the end of the page). For this reason, Bird proposed a new layout presenting the variant readings at the same time without resorting to the critical apparatus. The purpose is to print multiple versions of the text in parallel, creating a "multitext edition of Homer, one that would be expected not only to report variant readings but also to relate them as possible to different periods of history".[35] The weakness of this position is, in my opinion, the unproven equivalence of the aedic phase with the rhapsodic and then the Ptolemaic stage of the Homeric tradition – apart from a not entirely convincing and forced analogy between the Homeric tradition and that of the *New Testament*. Not all the variant readings have a certain tradition going back to the classical age as the 'Zenodotean' (and Aeschylean) δαῖτα in *Il.* I 5.[36] Ptolemaic papyri seem rather to hand down rhapsodic variants,[37] such as those transmitted from the medieval text of the *Homeric Hymns*. An informative example of the erroneous Bird's approach is his discussion[38] concerning the "minor textual variants" attested for *Il.* VI 287–8 in P.Sorb. inv. 2302 (TM 61240 = M-P³ 786.1 = LDAB 2380 = Allen-Sutton-West p480a): in his opinion, ἀόλιccαy κατά and θάλαμοy κατεβήcατο, rightly defined at first as "clear examples of spelling reflecting pronunciation",[39] "convey the memory of a live performance, with all its speed and dramatic intensity"; this spelling, continues Bird, "presumably would not happen if the lines were being dictated slowly and carefully". To affirm this, the phenomenon should be limited to the Homeric papyri, but it is notoriously

---

34 Bird 2010, 34–6; see also Dué – Ebbott 2010, 8: "where different written versions record different words, but each phrase or line is metrically and contextually sound, we must not necessarily consider one 'correct' or 'Homer' and the other a 'mistake' or an 'interpolation'".

35 Nagy 1996, 113. Already di Luzio 1969, 151 proposed to insert in the margin of the critical text the equivalent variations obtained from the papyri, so that the reader could choose the lesson according to his 'taste'; and the pleasure of choice would not be reserved only for the learned editor.

36 See Bird 2010, 34–7.

37 It is the case, probably, of *Il.* VIII 526, discussed by Bird 2010, 57–8, but see especially his n. 151.

38 Bird 2010, 92–6.

39 Bird 2010, 95.

widespread in texts of various nature, literary as well as documentary. A similar mis-interpretation concerns the spelling τόρ ῥα for τόν ῥα in *Il.* XVII 578, attested in P.Sorb. inv. 2303 (TM 61117 = M-P³ 948.2 = LDAB 2255 = Allen-Sutton-West p501c).[40]

Multitextuality has been also applied by DUÉ – EBBOTT 2010 to the tenth book of the *Iliad*:[41] the two scholars, in order to avoid presenting a critical text "that obscures the multiformity of the oral tradition", choose to print four witnesses that represent the state of the text in different historical periods (a papyrus of the II century BC, one of the III AD, one of the VI AD, and the *Venetus* A). The 'critical' text is followed by a commentary, which, although not inclusive of all the information of a critical appa-ratus, in the author's intentions makes it possible to better explore the differences between the witnesses as it is more focused on the chosen texts. In fact, this edition is the paper version of the digital project, and, as far as the paper edition is concerned, there are many reasons for perplexity: the main reason is that, although the authors wish ideally to reproduce any text that presents significant variants, they are forced, by necessity of space and time, to make a choice regarding the number of witnesses to be presented, and the difficulty increases due to the deliberate omission of the crit-ical apparatus. The *specimina* printed in the edition are certainly analysed in detail in the commentary and are cited together with other relevant witnesses, but, as ad-mitted by the editors themselves, without reaching the level of completeness and sys-tematicity that is instead of a critical apparatus. Moreover, such an edition is certainly no longer easily consultable with respect to a 'traditional' edition, above all because of the inconvenience of having to resort each time to the comment, separate from the text. The other disadvantages are: no translation is available, if not a few hints in the commentary (not even for the papyrus texts), a limited number of witnesses, no sup-plements for the papyrus' lacunae. Therefore, for this work the definition of 'critical edition' is not appropriate, given that the Dué and Ebbott reject the traditional oper-ation, as above anticipated ("we want to avoid presenting a critical text that obscures the multiformity of the oral tradition"). They believe that it is not the apparatus but the commentary that makes the edition 'critical', "but critical in a different way from what is usually indicated by the term".[42] As regards their definition of 'textual criti-cism', they consider that it is authentically exercised by not judging which text is right or wrong, but rather "to criticize what these texts contain in terms of the textual tra-dition and the oral tradition that preceded it"; in other words, one can only distin-guish what is a trivial scribal error from a genuine oral variation.[43] This unusual way of conceiving the critical edition, however, is the basis of an editorial product that, in

---

40 See BIRD 2010, 96–100.

41 See especially ch. 4, *Iliad 10. A Multitextual Approach* (pp. 153–66), and parts II and III (*Texts* and *Commentary*, pp. 169–382).

42 DUÉ – EBBOTT 2010, 153.

43 DUÉ – EBBOTT 2009, 7: "our textual criticism of Homeric epic, then, needs to distinguish what may genuinely be copying mistake and what are performance multiforms".

addition to being extremely uncomfortable in the consultation, in my opinion fails precisely in its primary purpose, which is to allow 'independent use' by the reader of these versions. The editors justify their unconventional editorial choice – which in fact is a 'non-editorial' choice, as they suspend any judgment on the value of the variant readings – claiming to join "to the goals of the Homer Multitext" Project, that is the realization of a digital scholarly edition of Homer.

The editors did not hesitate to describe the project as a revolution. The *Homer Multitext Project* stems from the belief that it cannot in any way apply the traditional editing system to works such as the *Iliad* and the *Odyssey*, because they originated from a long oral tradition without the aid of writing:

> in such a tradition in which the composition is occurring in the course of performance, there is no one "author of the original composition" to try to recover, for there is not only one composition, but also no other author.

It is the oral nature of the so-called Homeric poems that renders traditional editorial methods ineffective: the textual criticism that must be applied to them should not move in the direction of a 'paradigmatic' (or 'canonic') critical text, but should maintain "the integrity of each witness to allow for continual and dynamic comparison".[44] According to the editors' view, another traditional concept that must be abandoned is also that of "variant (reading)", because it implies a judgment on the quality of the variants presented by the manuscripts: we should therefore adopt the more neutral term "multiform (reading)". The logical consequence of this statement is that, if each variant reading is potentially 'authentic', the traditional critical apparatus no longer has any reason to exist: an approach of this kind "deliberately puts some central concepts and issues of conventional textual scholarship in crisis. The basic text, the text, the textual apparatus, and the variant".[45] Dué and Ebbott state with very explicit and categorical words: "the digital Multitext must be fundamentally different from these print editions in conception, structure and interface".[46] One point that is very relevant in the Project is the need to present the digitized text of all the witnesses but without the critical apparatus. Despite the conspicuous attention attracted by the *Homer Multitext Project*, the impression is that, beneath the statements of principle, there is currently very little concrete to compare with. In other words, there is a contrast between the limpidity and the firm certainty with which the Project guidelines have been presented and the indeterminacy with which practical questions are dealt with: e.g., how the texts of the manuscripts will be presented in the absence of a text-reference base? Which tool will be used to replace the traditional critical apparatus? In the section dedicated to the description of the project on the website of the *Homer Multitext Project*

---

44 Dué – Ebbott 2009, 5.
45 Vanhoutte 2007, 165–6.
46 Dué – Ebbott 2009, 2.

it still remains said that "unlike printed editions […] the Homer Multitext offers the tools for reconstructing a variety of times and places", without specifying what tools are involved and how they work. Moreover, the absence of precise guidelines on the structure of the Project had been admitted also by Dué and Ebbott,[47] who however had set aside the issue, considering it a detail ("but no matter what the details end up being, we have committed to three foundational principles: collaboration, open access, and interoperability").[48] It must have been this vagueness in structural terms, together with a certain insistence on the importance of leaving the final choice to the reader, to have aroused the lucid judgment of M.L. West: "Nagy seems to think that an editor should simply marshal the evidence in a non-committal way". While Nagy seems to assume that the Homeric editor's most important task is to suspend his own judgment to prevent it from undermining his reader's, West defends the right to actively exercise his own, concluding that they evidently have "very different concepts of the editor's role".[49] In Dué's and Ebbott's edition, the text of the 10th book of the *Iliad* becomes therefore four different Iliadic texts, three ancient and partial,[50] the fourth, medieval and complete (*Venetus* A), each witness of the Homeric perennial multitextuality. Despite the absolute peculiarity of the Homeric tradition, a methodology of this type leads not only to the renunciation of the critical text, as said, but also to the renunciation of a critical approach *tout court*: it is revealing the treatment of the variant (or better, multiform) readings, each of which seems *a priori* to testify to an aedic (re)composition-in-performance, even though the editors tend to overlook their possible, often probable un-aedic origin (rhapsodic variants, glosses, conjectures).[51]

5 Conclusions

As in many other scientific fields, the implementation of IT systems can lead to a methodological renewal only after a dissemination of their scientific use and only after having produced results superior to those obtained by past philologists with the traditional methods. This scenario could be achieved through better quantitative and

---

47 See also DUÉ – EBBOTT 2009, 33: "as we continue to build up our collection of texts, there are still questions to be answered about how to construct the architecture to achieve the visual representation we envision and that will achieve the results we have described here".

48 Three magic words in the digital editing. Their relevance is out of the question, but often editors do not go beyond the simple enunciation.

49 WEST 2001, 160 n. 5.

50 P.Mich. inv. 6972 (TM 61210 = M-P³ 864.1 = LDAB 2350 = Allen-Sutton-West p609); P.Berol. inv. 11911 + 17038 + 17048 + 21155 (TM 60757 = M-P³ 855.1 = LDAB 1883 = Allen-Sutton-West p425 + 430); P.Cair.Masp. inv. 67172-4 + P.Berol. inv. 10570 + P.Strasb. inv. G 1654 + P.Rein. II 20 (TM 61072 = M-P³ 658 = LDAB 2209 = Allen-Sutton-West p46).

51 So, e.g., the commentary to *Il.* X 10 (p. 247).

qualitative information processing, especially with reference to the collation of the manuscripts and to the stemmatic contamination. The availability of IT tools will increasingly allow to broaden and better evaluate the data of the manuscript tradition and could avoid an excessive limitation of the witnesses only for the need to reduce their number. The greater 'capacity' of the computer systems certainly allows a simpler and more efficient collection of the variant readings, sometimes useless for the constitution of the text but often valuable to study the textual tradition in a wider cultural sense. Regardless of the case of the *Homer Multitext Project*, there is however the sensation of an overestimation of those variant readings not really significant for the constitution of the text, overestimation which is equally pernicious compared to their scarce eight-twentieth-century consideration. Sometimes, this 'hypervalutation' seems to be promoted by some digital editors after having observed the limited or null ecdotic progress of the scholarly work. There are undeniable advantages in the digital scholarly editions, that have been clearly illustrated, for example, by Mastronarde (see above*)*, but it is also true, and obvious, that the ecdotic improvement derived from a more careful study of the paradosis is not necessarily produced by computer tools.[52] There will certainly be a moment when the critical editor will be supported or even replaced by the AI, but the task of choosing the best possible text in a methodologically correct manner cannot but remain the essential purpose of the philological activity for most of the manuscript traditions handed down by a plurality of manuscripts.

6 Bibliography

Babeu, A. (2011), *"Rome Wasn't Digitized in a Day". Building a Cyberinfrastructure for Digital Classicists*, Washington (DC), URL: http://www.clir.org/pubs/abstract/pub150abst.html.

Bernagou, E. – Palumbo, G. – Rinoldi, P. (2016), *L'informatica al servizio dell'ecdotica: l'edizione della Chanson d'Aspremont*, "Le forme e la storia. Rivista di Filologia Moderna" 9(1), 39–62.

Berti, M. – Almas, B. – Crane, G.R. (2016), *The Leipzig Open Fragmentary Texts Series (LOFTS)*, in *Digital Methods and Classical Studies*, ed. by N.W. Bernstein and N. Coffee [DHQ Themed Issue 10(2)], URL: http://www.digitalhumanities.org/dhq/vol/10/2/000245/000245.html.

Berti, M. – Blackwell, C. – Daniels, M. – Strickland, S. – Vincent-Dobbins, K. (2016), *Documenting Homeric Text-Reuse in the* Deipnosophistae *of Athenaeus of Naucratis*, in *Digital Approaches and the Ancient World*, ed. by G. Bodard, Y. Broux, and S. Tarte [BICS Themed Issue 59(2)], 121–39, URL: http://onlinelibrary.wiley.com/doi/10.1111/j.2041-5370.2016.12042.x/full.

Berti, M. – Daniels, M. – Strickland, S. – Vincent-Dobbins, K. (2016), *Modelling Taxonomies of Text Reuse in the* Deipnosophists *of Athenaeus of Naucratis: Declarative Digital Scholarship*, in *Digital Humanities 2016: Conference Abstracts*, Kraków, 135–7, URL: http://dh2016.adho.org/abstracts/46.

---

52 E.g., the examples drawn from the Platonic paradosis by Cufalo – Muggittu 2016, 92, in order to illustrate the relevance of a more comprehensive *recensio*, have nothing to do with the IT tools.

BIRD, G.D. (2010), *Multitextuality in the Homeric Iliad. The Witness of the Ptolemaic Papyri*, Cambridge (MA).

BOSCHETTI, F. (2009), *A Corpus-based Approach to Philological Issues*, PhD Diss., Trento.

CUFALO, D. – MUGGITTU, V. (2016), *"Digital Native" Critical Editions and Homemade School Text Analysis: The Hyper Project*, "Literatūra" 58(3), 88–107.

DE FAVERI, L. (2002), *Die metrischen Trikliniusscholien zur byzantinischen Trias des Euripides*, Stuttgart.

DI LUZIO, A. (1969), *I papiri omerici d'epoca tolemaica e la costituzione del testo dell'epica arcaica*, RCCM 11, 3–151.

D'IORIO, P. (2010), *Qu'est-ce qu'une édition génétique numérique?*, "Genesis (Manuscrits – Recherche – Invention)" 30, 49–53, URL: https://genesis.revues.org/116.

DUÉ, C. – EBBOTT, M. (2009), *Digital Criticism: Editorial Standards for the Homer Multitext*, DHQ 3(1), URL: http://www.digitalhumanities.org/dhq/vol/003/1/000029/000029.html.

DUÉ, C. – EBBOTT, M. (2010), *Iliad 10 and the Poetics of Ambush. A Multitext Edition with Essays and Commentary*, Washington (DC).

FRANZINI, G. (2012–), *Catalogue of Digital Editions*, URL: https://github.com/gfranzini/digEds_cat.

FRANZINI, G. – ANDORFER, P. – ZAYTSEVA, K. (2016–), *Catalogue of Digital Editions: The Web Application*, URL: https://dig-ed-cat.acdh.oeaw.ac.at.

FROGER, J. (1968), *La critique des textes et son automatisation*, Paris.

HAINSWORTH, J.B. (1982), ed., *Omero. Odissea, II (ll. V–VIII)*, Milano.

HARRIS, N. (2014), *Col piede sbagliato, e con i piedi di piombo*, "Ecdotica" 11, 73–85.

ITALIA, P. – TOMASI, F. (2014), *Filologia digitale. Fra teoria, metodologia e tecnica*, "Ecdotica" 11, 112–30.

KAIBEL, G. (1887–90), cur., *Athenaeus. Deipnosophistae*, I–III, Lipsiae.

MALASPINA, E. – DELLA CALCE, E. (2017), *Classici e computer: verso la transdisciplinarità?*, in *Humanities e altre scienze. Superare la disciplinarità*, a c. di M. Cini, Roma, 49–65.

MARTINI, G. (2013), *I papiri tolemaici dell'Iliade*, Diss. Parma.

MCNAMEE, K. (2007), *Annotations in Greek and Latin Texts from Egypt*, Ann Arbor (MI).

MONELLA, P. (2012), *Why Are There No Digital Scholarly Editions of 'Classical' Texts?*, in *Constitutio textus: la 'ricostruzione' del testo critico / Constitutio Textus: Establishing the Critical Text. IV Incontro di Filologia digitale (Verona, 13–15 September 2012)*, Verona (pre-print draft), URL: http://www1.unipa.it/paolo.monella/lincei/files/why/why_paper.pdf.

NAGY, G. (1996), *Poetry as Performance*, Cambridge.

PIERAZZO, E. (2011), *A Rationale of Digital Documentary Editions*, "Literary and Linguistic Computing" 26(4), 463–77.

PIERAZZO, E. (2015), *Digital Scholarly Editing. Theories, Models and Methods*, Farnham – Burlington (VT).

SCHMIDT, D. (2010), *The Inadequacy of Embedded Markup for Cultural Heritage Texts*, "Literary and Linguistic Computing Advance Access", April 16, 1–20, URL: http://hki.uni-koeln.de/archive/hki2016/sites/all/files/courses/3227/Schmidt-2010.pdf.

SCHMIDT, D. (2016), *Ecdosis: Scholarly Editions for the Web*, in *Edizioni Critiche Digitali / Digital Critical Editions. Edizioni a confronto / Comparing Editions*, a c. di P. Italia e C. Bonsi, Roma, 93–103.

SCHWARTZ, E. (1887–91), cur., *Scholia in Euripidem*, I–II, Berolini.

VANHOUTTE, E. (2007), *Traditional Editorial Standard and the Digital Edition*, in *Learned Love. Proceedings of the Emblem Project, Utrecht Conference on Dutch Love Emblems and the Internet*, ed. by E. Stronks and P. Boot, The Hague, 157–74, URL: http://emblems.let.uu.nl/static/images/project/learned_love_157-174.pdf.

WEST, S. (1967), *The Ptolemaic Papyri of Homer*, Köln – Opladen.

WEST, M.L. (1973), *Textual Criticism and Editorial Technique*, Stuttgart.

WEST, M.L. (2001a), *Studies in the Text and Transmission of the* Iliad. *Disquisition and Analytical Commentary*, München – Leipzig.

WEST, M.L. (2001b), *West on Nardelli and Nagy on West*, BMCR 2001.09.06, URL: http://bmcr.bryn-mawr.edu/2001/2001-09-06.html.
WEST, M.L. (2017), cur, *Homerus. Odyssea*, Berlin – Boston.

Part 2: **Linguistic Perspectives**

Marja Vierros
Linguistic Annotation of the Digital Papyrological Corpus: *Sematia*

1 Introduction: Why to annotate papyri linguistically?

Linguists who study historical languages usually find the methods of corpus linguistics exceptionally helpful. When the intuitions of native speakers are lacking, as is the case for historical languages, the corpora provide researchers with materials that replaces the intuitions on which the researchers of modern languages can rely. Using large corpora and computers to count and retrieve information also provides empirical back-up from actual language usage. In the case of ancient Greek, the corpus of literary texts (e.g. *Thesaurus Linguae Graecae* or the Greek and Roman Collection in the *Perseus Digital Library*) gives information on the Greek language as it was used in lyric poetry, epic, drama, and prose writing; all these literary genres had some artistic aims and therefore do not always describe language as it was used in normal communication. Ancient written texts rarely reflect the everyday language use, let alone speech. However, the corpus of documentary papyri gets close. The writers of the papyri vary between professionally trained scribes and some individuals who had only rudimentary writing skills. The text types also vary from official decrees and orders to small notes and receipts. What they have in common, though, is that they have been written for a specific, current need instead of trying to impress a specific audience. Documentary papyri represent everyday texts, utilitarian prose,[1] and in that respect, they provide us a very valuable source of language actually used by common people in everyday circumstances.

This significant text corpus is openly available to us in digital form. The *Papyrological Navigator* (PN)[2] hosts the *Duke Databank of Documentary Papyri* and provides a search engine as well. However, any deeper linguistic research cannot be performed. The search engine at PN is mainly designed for the needs of historians and editors of papyrus texts in locating parallels and sources using word-string searches. In order to utilize the text corpus linguistically, it needs to be enriched with linguistic information, i.e. it needs to be linguistically annotated.[3] Linguistic annotation can concern many different levels of language, usually morphology, syntax, semantics,

---

1 Cf. Wagner – Outhwaite – Beinhoff 2013, 4.

2 http://papyri.info/.

3 A very clear textbook on linguistic annotation and corpus linguistics in general is Kübler – Zinmeister 2015. See also e.g. Wynne 2005 on developing linguistic corpora.

or pragmatics. Even the basic morphological annotation alone can provide for complex linguistic queries. The literary Greek corpus has recently been automatically lemmatized and morphologically parsed.[4] Greek that is found in papyri deserves to be similarly treated so that the literary language can be compared with the utilitarian prose found in papyri, enabling our views on historical developments and variation of Greek language to be as full as they can be.

In this paper, I will discuss the criteria and approach which I have chosen while planning the *Sematia* corpus and platform.[5] While this is an ongoing process and plans often are subject to change, it is still worthwhile to explain what lies behind the selected approach, what the future plans are and possible new directions and, finally, what can be achieved with all this work.

2 Corpus design

One key factor in corpus design generally is that the corpus is representative. Whether we want a holistic or strictly selected corpus, depends on the research questions for which the corpus is meant to provide answers. If we want answers from a certain domain of texts (e.g. private letters), we select only those texts into the corpus. Similarly, whether we want a synchronic or diachronic corpus depends on whether we want to examine changes in language used within a certain time span or not. In historical linguistics, corpora are generally diachronic.

The papyrological corpus in PN is a growing and a changing one. It includes all published documentary papyri, and the Greek material ranges approximately from the IV century BC to the IX century AD. Newly published texts are added into the database by the academic community of papyrologists via the online *Papyrological Editor* (PE), where a board checks and votes on the submissions.[6] Also, mistakes (typos or wrong readings etc.) in the texts that already exist in the corpus, can be corrected via the same *Editor*. This is one reason for the idea that *Sematia* should also be kept open-ended, so that ideally it could include the whole corpus, which represents the Greek used in documentary papyri for a period of about a thousand years. Thus, at the moment, the corpus design is a loose one, but users (both the annotators and the researchers who only wish to perform queries) can decide on a case-by-case basis what they want to annotate or include in their searches. Once a version of a text has been annotated, that annotation is stable, but if the system alarms us that there has been a change introduced into the base text in the PN, the annotator (or someone else,

4 Celano 2017.

5 https://sematia.hum.helsinki.fi. I warmly thank the developer, Erik Henriksson, for all his ideas and efforts.

6 Sosin 2010, cf. also Reggiani 2017, 232–40.

for that matter) can renew the annotation on that text, if it seems warranted.[7] The process of getting texts annotated is slow at the moment, since it is performed semi-automatically (more on this aspect below). The choice of texts to be annotated is not authoritatively dictated by us; the choice is made by the users, so anyone wanting to have a specifically chosen set of material, can proceed in annotating the papyri. This way s/he also makes a contribution towards the annotation of the whole corpus. And when there are more texts already annotated, each researcher may select his/her own subcorpus and perform queries only on them (either in the *Sematia* platform or after downloading all the selected annotations for external use). The latter option makes the research more easily replicable (a basic requirement in corpus linguistic research).

Corpus design also includes deciding over the level of annotation and what features are annotated and how. At the moment, our basic approach is to include the morphological and syntactic annotation in the form of dependency treebanks. We follow the *Ancient Greek and Latin Dependency Treebank* system.[8] *Sematia* is designed to provide a 'basic' level of annotation, because we have this holistic idea of the whole corpus eventually being annotated; the research questions must not in this case be strictly decided beforehand. However, since the automatic morphological parsing has been performed on literary texts as mentioned above, this is a logical next step for the whole papyrological text corpus as well. This, in turn, would make the manual syntactic treebanking somewhat quicker, as the morphological forms would be more accurate than they are now to begin with (on the process of annotation, more detailed description below).

3 How to annotate papyri?

Why should we devote a section on how to linguistically annotate papyrus texts? Because the papyri represent ancient textual material often preserved in a fragmentary condition. The organic writing material has suffered damage of many kinds. But, due to the importance of papyri as a source, papyrologists work very hard on reading, transcribing, and reconstructing them, i.e. editing the text, so that other researchers can also use that source. Still, many gaps and question marks can remain in the editions. All this is encoded within the text in the digital edition, in TEI EpiDoc XML,[9] and for this reason we do not have simple access to the raw text that could simply be uploaded for some linguistic annotation tool. In fact, the editorial work gives us

---

7 This type of alarm system has not yet been established, but it is on our agenda.

8 https://perseusdl.github.io/treebank_data; BAMMAN – CRANE 2011.

9 https://sourceforge.net/p/epidoc/wiki/Home.

plenty of material that we can and should also use in the linguistically annotated corpus. Therefore, we need to preprocess the texts available in the PN in a certain way.

3.1 Preprocessing

The *Sematia* tool was first developed mainly for the above-mentioned preprocessing need. It creates two parallel layers of the same text; one being a sort of diplomatic edition (called "original"), and the other including the editorial suggestions (called "standard"). The tool has already been described in another article,[10] thus I will not present the details here. What makes the *Sematia* corpus special, is that *both* of these parallel layers are linguistically annotated. This way it is possible to study only the version that has truly been preserved for us (the original layer), or to compare the actual preserved text with its standardized version. The differences in this comparison can be turned into a third layer (called "variation"), which I will briefly discuss later.

3.2 Annotation

In order for this corpus to be beneficial for all Greek linguists, I decided that we should follow the same scheme and standard used in the corpora of ancient Greek. This means the *Ancient Greek and Latin Dependency Treebank* that includes Greek literature. In addition, the PROIEL treebank (New Testament and some Greek prose) follows the Dependency Grammar.[11] In the annotation of papyri, we follow the Guidelines of AGDT.[12] At the moment, we use the external annotation environment, *Arethusa*, provided by the *Perseids* Platform,[13] with which we have an API integration in *Sematia*. This means that a text can be exported directly from *Sematia* to *Perseids* and *Arethusa*, and after it has been annotated, a member of the *Sematia* board (at the moment the project director) goes through the annotations in *Perseids* and either accepts or returns them to the annotator to be corrected. After the approval, the treebanks are committed back to *Sematia* (both the *GitHub* repository and the online site).

The process of annotation in *Arethusa* includes the tokenization; i.e. tokenization is done when the plain text is imported into *Arethusa*, not into *Sematia*. The text receives an automatic lemmatization and morphological tagging (by *Morpheus*). But all

---

10 Vierros – Henriksson 2017.

11 Both treebanks have also been modified for the Universal Dependencies site (http://universaldependencies.org), where they can be accessed together with many other languages.

12 Version 1.1: Bamman – Crane 2008, version 2.0: Celano 2014. Version 2.0 is to be followed, but version 1.1 has sometimes more useful examples and more detailed explanations.

13 http://sites.tufts.edu/perseids.

the lemmas and morphological tags need to be checked and corrected by the human annotator; there are several forms and lemmas in the papyri, which *Morpheus*, being designed for classical Greek, does not recognize, for example the Egyptian names. Moreover, *Morpheus* does not do well in selecting the correct form from several homonyms. The syntactic annotation and dependencies have to be performed manually by the annotator. In other words, using *Arethusa* is convenient up to a point; but it is also quite laborious and thus expensive as it needs human resources: skilled annotators and their time. Nevertheless, in the end, we do get accurate annotations that can most likely be used in training automatic syntactic and morphological parsers in the future.

The process can be presented by an example with images. Our sample sentence is the second sentence of a letter from Petenephotes to Valerius, written on a potsherd in the garrison of Mons Claudianus in the Eastern Desert (O.Claud. II 245,2–7; mid II century AD):

[1] [καλῶς] |³ πυήσις, ἄδελφε, ἐὰ[ν ἔλθῃ] |⁴ ἡ πορήε τῇ νυκτὶ ταύτῃ ⸌πέμψας μοι / |⁵ τρία ζεύγη ἄρτων ἐπὶ οὐκ ἐ|⁶χο ἄρτους καὶ ὅταν ἔλθῃ ἡ πο|⁷ρήα πέμψω συ αὐτά.

3. *l.* ποιήσεις   4. *l.* πορεία   5–6. *l.* ἔ|χω   6–7. *l.* πο|ρεία   7. *l.* σοι

Please, brother, if the caravan arrives tonight, send me three pairs of bread as I do not have any bread and when the caravan arrives I shall send them to you.

Note that the apparatus has several corrections (standardizations), but not for the ι/ει confusion in the conjunction ἐπὶ (l. ἐπεί), l. 5. This is the standard practice in this edition. Other so-called orthographic mistakes are usually standardized in the apparatus, but not the most common one between ι / ει, because the editors apparently consider this such a common, parallel variant that it can no longer be considered as a 'mistake' (see also the chapter by J. Stolk in this volume for problems that this type of fluidity between editorial corrections can cause).

The standard and the original layers of this sentence in the *Arethusa* treebank tool are presented in Figg. 1 and 2. Only the syntactic trees can be seen in the screenshots, and only one lemma/morphological analysis (that of the highlighted word), in this case the conjunction mentioned above. This is emphasized here, because an automatic parser would automatically take this word as the preposition ἐπὶ, but when the human annotator checks the sentence, s/he notices that the preposition is not the correct interpretation, and can make the necessary correction, even though the word is not editorially corrected in the original electronic source of ours, in the PN.

The differences between the layers are apparent in the images; the supplied text, for example, is not annotated in the original layer, it is represented with a dummy marker SU so that the annotator notices that something is missing there and the supplemented word does not end up in the corpus of original layers. This also leaves some of the branches of the sentence tree hanging in the air, as some words that

would be the heads on which other words depend on, are not preserved in the papyrus. The non-standard orthography in the original will not prevent the annotator from recognizing and marking correct lemmata for the forms, thus lemma searches will find all variant spellings of the words from the original layers.
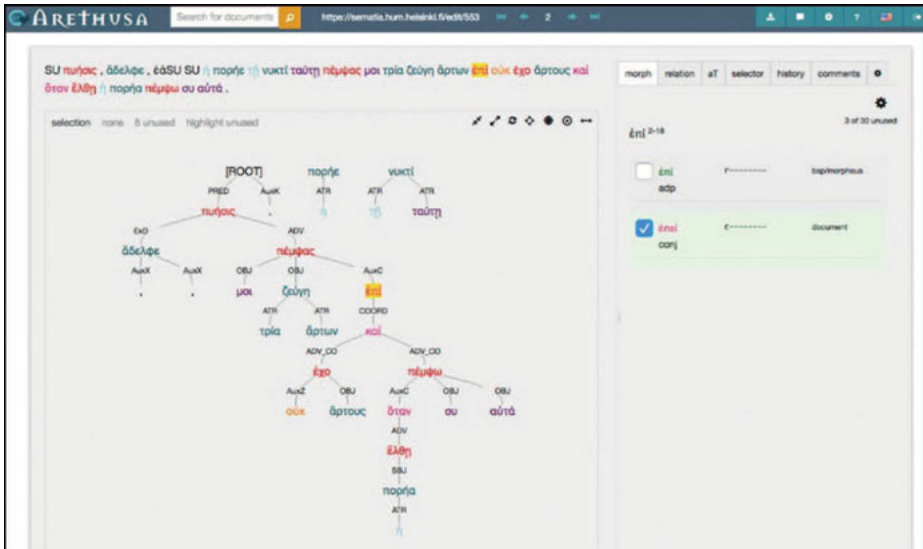
Fig. 1: Original layer of the sentence [1] in *Arethusa*.
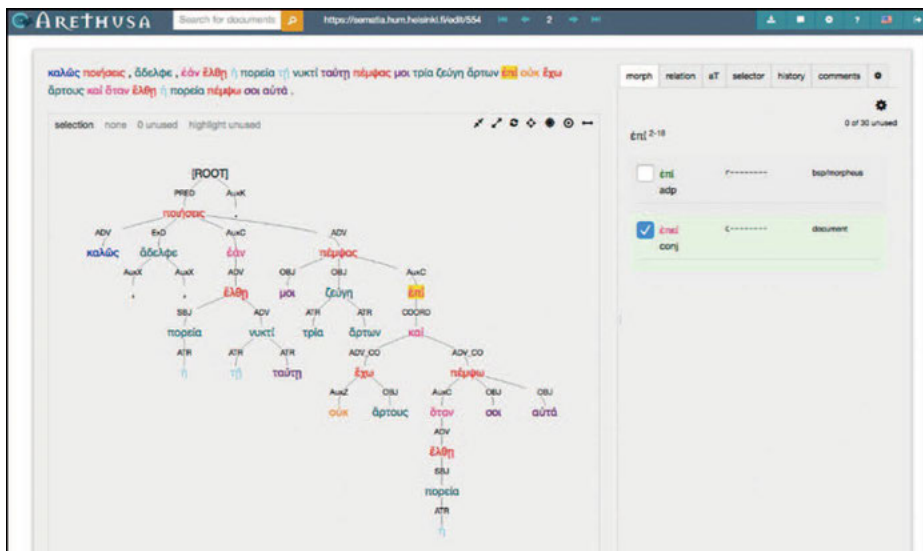
Fig. 2: Standard layer of the sentence [1] in *Arethusa*.

The underlying XML forms show us what the whole annotation entails. Fig. 3 presents the XML of the original layer's annotation of the same sentence [1].

```xml
<sentence document_id="https://sematia.hum.helsinki.fi/edit/553" id="2" span="" subdoc="">
  <word form="SU" head="" id="1" lemma="" postag="" relation="" />
  <word form="πυήσις" head="0" id="2" lemma="ποιέω" postag="v2sfia---" relation="PRED" />
  <word form="," head="4" id="3" lemma="punc1" postag="u--------" relation="AuxX" />
  <word form="ἄδελφε" head="2" id="4" lemma="ἀδελφός" postag="n-s---mv-" relation="ExD" />
  <word form="," head="4" id="5" lemma="punc1" postag="u--------" relation="AuxX" />
  <word form="ἐὰSU" head="" id="6" lemma="" postag="" relation="" />
  <word form="SU" head="" id="7" lemma="" postag="" relation="" />
  <word form="ἡ" head="9" id="8" lemma="ὁ" postag="l-s---fn-" relation="ATR" />
  <word form="πορήε" head="" id="9" lemma="πορεία" postag="n-s---fn-" relation="" />
  <word form="τῇ" head="11" id="10" lemma="ὁ" postag="l-s---fd-" relation="ATR" />
  <word form="νυκτὶ" head="" id="11" lemma="νύξ" postag="n-s---fd-" relation="" />
  <word form="ταύτη" head="11" id="12" lemma="οὗτος" postag="p-s---fd-" relation="ATR" />
  <word form="πέμψας" head="2" id="13" lemma="πέμπω" postag="v-sapamn-" relation="ADV" />
  <word form="μοι" head="13" id="14" lemma="ἐγώ" postag="p1s---md-" relation="OBJ" />
  <word form="τρία" head="16" id="15" lemma="τρία" postag="n-p---na-" relation="ATR" />
  <word form="ζεύγη" head="13" id="16" lemma="ζεῦγος" postag="n-p---na-" relation="OBJ" />
  <word form="ἄρτων" head="16" id="17" lemma="ἄρτος" postag="n-p---mg-" relation="ATR" />
  <word form="ἐπὶ" head="13" id="18" lemma="ἐπεί" postag="c--------" relation="AuxC" />
  <word form="οὐκ" head="20" id="19" lemma="οὐ" postag="d--------" relation="AuxZ" />
  <word form="ἔχο" head="22" id="20" lemma="ἔχω" postag="v1spia---" relation="ADV_CO" />
  <word form="ἄρτους" head="20" id="21" lemma="ἄρτος" postag="n-p---ma-" relation="OBJ" />
  <word form="καὶ" head="18" id="22" lemma="καί" postag="c--------" relation="COORD" />
  <word form="ὅταν" head="27" id="23" lemma="ὅταν" postag="c--------" relation="AuxC" />
  <word form="ἔλθη" head="23" id="24" lemma="ἔρχομαι" postag="v3sasa---" relation="ADV" />
  <word form="ἡ" head="26" id="25" lemma="ὁ" postag="l-s---fn-" relation="ATR" />
  <word form="πορήα" head="24" id="26" lemma="πορεία" postag="n-s---fn-" relation="SBJ" />
  <word form="πέμψω" head="22" id="27" lemma="πέμπω" postag="v1sfia---" relation="ADV_CO" />
  <word form="συ" head="27" id="28" lemma="σύ" postag="p-s---mn-" relation="OBJ" />
  <word form="αὐτά" head="27" id="29" lemma="αὐτός" postag="p-p---na-" relation="OBJ" />
  <word form="." head="0" id="30" lemma="punc1" postag="u--------" relation="AuxK" />
</sentence>
```

Fig. 3: XML of the treebanked original layer of the sentence [1].

We can see that the annotation includes the existing form of the word in the sentence, its head, its lemma, the postag and the syntactic relation of the word in the sentence. The postag includes the whole morphological analysis: part-of-speech, person, number, tense, mood, voice, gender, case and degree. For example, the form πεμψας (word id 13) is a verb, singular, aorist participle active in the masculine nominative. The postag gives the very basic morphological analysis, and we could occasionally hope for something more specific, such as distinguishing proper nouns from common nouns or possessive pronouns from other pronouns, but as of this moment, *Morpheus* gives us these. In the future, other automatic parsers might take these distinctions into account more easily. However, even this morphological analysis enables us to search complex linguistic structures, especially when combined with the lemma and syntactic annotations. This, I think, is sufficient to fulfill the need of basic linguistic annotation for the *Sematia* corpus. Other levels of annotation, e.g. semantic or information structure annotations, would take considerably more time and effort.

4 Metadata and its purposes

The date and place of origin of each text are vital when we wish to see in which time periods and in which areas certain linguistic features appear. They are generally provided in the papyrus editions and presented also in the PN metadata field, from whence they are automatically drawn into *Sematia*.

As mentioned already in VIERROS – HENRIKSSON 2017, we add some metadata, which is not available in PN, namely aspects relating to the handwriting and the writers vs. authors. Some changes have been planned for these metadata fields and they will be implemented in the near future. The purpose is to identify text parts written in the same hand. When imported to *Sematia*, each document is divided into 'acts of writing' by the element `<handShift>`, i.e. each section written by a different writer receives its own layers and treebanks. Since there are often papyrus archives in which the same hand can have written several documents, it is important to link these acts of writing together, so that we can also try to study idiolects and compare certain writers to others. At the moment of writing, we can add metadata concerning the handwriting[14] and concerning the writer, the author and the addressee.[15] See Fig. 4 for an example on the metadata in O.Claud. II 245 (which only has one hand). In many cases, however, the name of the actual writer is not known, e.g. in private letters the sender of the letter is taken as the author, but the actual writer is not necessarily the same person as the author, nor is he named. In contracts, the names of the contracting parties are mentioned, but the scribe who draws up the text or who pens down the letters onto the papyrus often remains unnamed. Therefore, the *Trismegistos People* ID cannot be used in identifying the hands, since we have so many hands without names to connect them with. Our intention is to give each hand an ID of its own. The hands that have been identified to come from one writer (sometimes a very difficult task), can be connected to the same ID. The hand-ID will make the current metadata field "Same hand" obsolete.[16]

For the purposes of studying linguistic register and features typical of certain text types, we have also included the fields in which we can insert metadata on the text type and the addressee.

---

14 There are fields for the description of the handwriting in the edition or some other scholarly source, the description of the handwriting by the annotator, and the "Same hand" field, i.e. list of other documents, where the same hand is said to appear. These fields are text-based, and thus they do not provide good searchable data. Every papyrologist is also well aware of the lack of precision of these descriptions in different editions.

15 For each person the annotator can add the name, title and the *Trismegistos People* ID (http://www.trismegistos.org/ref/index.php).

16 In its current state, the field is not very usable, user-friendly or accurate; the list of other documents where the same hand appears is done in stable URLs of the documents in PN, but one document can contain several hands.

Fig. 4: Screenshot of the main view from *Sematia*, when the document O.Claud. II 245 is expanded (but O.Claud. II 243 and 246 are not). On the right, the metadata inserted in *Sematia* by the annotator is visible. The field "Same hand" is extensive with many documents also written in Petenephotes's hand. The editor mentions that the writer is Petenephotes himself,[17] thus he is both the author and the writer. Clicking from the blue "original" or green "standard" buttons would take you to the text, and clicking the paper icon next to those buttons, you could view the treebank XML.

5 Sample results, i.e. what queries can find

The treebank XML files (including the metadata) in *Sematia* can be exported for querying in external treebank query tools.[18] It is possible to export the treebanks of all layers together, or choose the original or standard layers separately. I will not go through all the possibilities the external search engines can give for linguists;[19] I will describe some sample searches that can be performed on the *Sematia* site itself.[20] There, too, it is possible to search only from the treebanks of the original layers or only from the standard layers, but one of the essential features is the possibility to find instances where the original and standard layers differ. This is where we can get

---

17 Bülow-Jacobsen 1997, 69.

18 E.g. *SETS Treebank Search*, *PML Tree Query Engine* or *XQuery/BaseX*, cf. Vierros – Henriksson 2017, 13.

19 One thorough treebank-based study on ancient languages is Korkiakangas 2016, in which the author has been able to study under which conditions the Latin accusative began to be used as the subject case in VIII and IX centuries.

20 https://sematia.hum.helsinki.fi/tools.

more deeply into linguistic variation. For example, it is very simple to search for instances where one grammatical case is used when editors have thought that a different case would have been more understandable, or more standard (what the editorial standardizations might have meant in different times when papyri have been edited, see the chapter by J. Stolk in this volume). The search fields in *Sematia* employ Regular Expressions (regex). The searches can naturally be limited in multiple ways, either by metadata fields or by the other field related to linguistic annotation, e.g. searching only objects or subjects, or only verbs or pronouns. More complex searches combining several words or forms would need to be made externally.

An example search concerning the grammatical case, the dative instead of the genitive, is presented in Fig. 5. Since the postag holds the case in the 8th place of the string, we can use the values for dative (d) in the original layer's postag field and genitive (g) in the 8th place in the standard layer's field, and let other places of the string be whatever else by using the wildcard (.); the beginning of the string is marked by (^). The values (d) and (g) can have different meanings in other positions in the postag, thus it is good to define the exact location. In other words, when using the search, it is vital to know how the annotations have been made, i.e. what each symbol means e.g. in the postag field. The guidelines of annotation need to be known and understood.

Fig. 5: A screenshot of the search and results in *Sematia* for the dative in the original layer vs. the genitive in the standard layer.

The search gives eight results with the limited data we have in *Sematia* at the moment (2017, ca. 100 papyri). The result list can be ordered according to different fields, in Fig. 5 it is ordered by the document name. We can see that some of the instances may, in fact, signal orthographic confusion based on phonological variation rather than case confusion (e.g. Νεχουτωι / Νεχουτου),[21] but some of the instances more clearly tell that the writer has, for some reason, really chosen the dative rather than the expected genitive (e.g. Μαρονατι / Μαρονατος). Similarly, we could bring up e.g. all prepositions in the texts by simple postag query (^r), or see where singular verb forms appear instead of plural verb form (^v.s vs. ^v.p). In the latter search, the results again point to the interplay of phonological factors confusing the morphological interpretations. See Fig. 6, where two out of three of the singular vs. plural verb form are forms consisting of graphemes αι / ε, both marking the phoneme /e/ at this time, and the third one has α / ε confusion, which was also perhaps due to weak pronunciation of the unstressed vowel. These results give us material for further research on phonology playing a part in the morphological mergers in Greek, and the impact of education in writers' ability or inability to use standard orthography in such occasions, but they also provide us with material for enhancing our tools in the future.

Fig. 6: A screenshot of the search in *Sematia* for a singular verb in the original layer vs. a plural verb in the standard layer.

---

21 See, however, DAHLGREN 2017, 90 ff. on phonological variation of /o, u/ possibly playing a role in case variation.

6 Future plans

A variation layer has been on our agenda since the beginning and it was discussed already in the previous article to some extent.[22] With the above described method of comparison between the original and standard layers, we can only find variation (instances where there really are differences between the layers), when there is a regularization in the PN, or when the annotator has marked these differences in the treebank XML after seeing a difference not available in the PN version. These comparisons and differences are planned to be automatically retrieved into *Sematia* to form the basis for the variation layer. In addition, we do need a way to manually encode other types of linguistic variation in this layer for several reasons. For example, there is a need to further specify certain differences as more phonological or more morphological in nature. Secondly, some variation is impossible to detect from the annotations when the postag does not really describe what we have in the text. I will give an example of this type of case with one sentence from a letter written by Ammonius to Apollonius (O.Claud. I 155,3–5; II century AD):

> [2] Ἁρπαήσιος ὁ κιβαριάτης εἴ|⁴ρηκέ μοι ὅτι ἐπιστολὴν ἔλα|⁵βα ἀπὸ τῆς γυναικός μου.
>
> *Harpaesius, the* cibariator*, has told me that I have got a letter from my wife.*

The form ἔλαβα, "I got", has not been corrected in the apparatus, even though it represents mixed morphology; the aorist of the verb λαμβάνω would be ἔλαβον according to the classical standard (the second i.e. 'strong' aorist), but in the *Koine* the athematic endings of the first i.e. 'weak' aorist (-α for the first person) were occasionally used (and they are the ones used in modern Greek).[23] In the Mons Claudianus *ostraka* so far annotated in *Sematia*, there are nine attestations of the form ἔλαβα (plus three times written as αἴλαβα),[24] but the editor has fluctuated in correcting it in the apparatus (see Fig. 7). We can find this word by using the word search, but as can be seen from the postag, it is not possible to indicate this type of variation there; the postag is the same in both ἔλαβα and ἔλαβον: first person singular aorist form. It would be very convenient to mark this up in the separate variation layer as mixed morphological endings in the aorist.

---

22 Vierros – Henriksson 2017, 13.

23 Cf. Horrocks 2010, 109–10 and 143–4 on the developments of past-tense morphology.

24 All three in O.Claud. II 236.

S-W	Word	Lemma	Relation	Postag	Hand	Doc.
3-5	ἔλαβα ἔλαβα	λαμβάνω λαμβάνω	PRED PRED	v1saia--- v1saia---	m1	o.claud.1.153
2-8	ἔλαβα ἔλαβα	λαμβάνω λαμβάνω	OBJ OBJ	v1saia--- v1saia---	m1	o.claud.1.155
2-4	ἔλαβα ἔλαβα	λαμβάνω λαμβάνω	PRED_CO PRED_CO	v1saia--- v1saia---	m1	o.claud.1.166
2-1	ἔλαβα ἔλαβα	λαμβάνω λαμβάνω	PRED PRED	v1saia--- v1saia---	m1	o.claud.1.167
6-14	ἔλαβα ἔλαβον	λαμβάνω λαμβάνω	OBJ_CO OBJ_CO	v1saia--- v1saia---	m1	o.claud.2.226
6-17	ἔλαβα ἔλαβον	λαμβάνω λαμβάνω	OBJ_CO OBJ_CO	v1saia--- v1saia---	m1	o.claud.2.226
8-3	ἔλαβα ἔλαβον	λαμβάνω λαμβάνω	PRED PRED	v1saia--- v1saia---	m1	o.claud.2.227
7-12	ἔλαβα ἔλαβον	λαμβάνω λαμβάνω	OBJ OBJ	v1saia--- v1saia---	m1	o.claud.2.228
2-6	ἔλαβα ἔλαβα	λαμβάνω λαμβάνω	OBJ_CO OBJ_CO	v1saia--- v1saia---	m1	o.claud.2.236

Fig. 7: A screenshot of the search results in *Sematia* for the word form 'ελαβα'. In the "word" column, the words in green come from the "original" layers and the words in black come from the "standard" layers. In O.Claud. volume I, the form was not standardised according to the classical norm, whereas in volume II it was (with one exception).

We will be developing *Sematia* and similar tools further.[25] One idea is to have the whole papyrological corpus already present in *Sematia*, and updated in set intervals, i.e. there would no longer be the need to import texts individually. Phonological searches will be enabled on the whole corpus. We also aim at developing an automatic morphological parser for Greek found in papyri, with more accurate analysis than what *Morpheus* currently has.

7 Bibliography

Bamman, D. – Crane, G. (2008), *Guidelines for the Syntactic Annotation of the Ancient Greek Dependency Treebank* 1.1*,* The Perseus Project, Tufts University, URL: http://nlp.perseus.tufts.edu/syntax/treebank/greekguidelines.pdf.

Bamman, D. – Crane, G. (2011), *The Ancient Greek and Latin Dependency Treebanks*, in *Language Technology for Cultural Heritage*, Berlin–Heidelberg, 79–98.

Bamman, D. – Mambrini, F. – Crane, G. (2009), *An Ownership Model of Annotation: The Ancient Greek Dependency Treebank,* in *Proceedings of the 8th Workshop on Treebanks and Linguistic Theories (TLT8),* at URL*:* http://www.perseus.tufts.edu/~ababeu/tlt8.pdf.

Bülow-Jacobsen, A. (1997), *The Correspondence of Petenephotes (243–254),* in *Mons Claudianus Ostraca Graeca et Latina II*, ed. by J. Bingen, A. Bülow-Jacobsen, W.E.H. Cockle, H. Cuvigny, F. Kayser, and W. Van Rengen, Le Caire.

Celano, G.G.A. (2014), *Guidelines for the Annotation of the Ancient Greek Dependency Treebank 2.0.*, URL: https://github.com/PerseusDL/treebank_data/edit/master/AGDT2/guidelines.

Celano, G.G.A. (2017), *Lemmatized Ancient Greek Texts*, URL: https://github.com/gcelano/LemmatizedAncientGreekXML.

Celano, G.G.A. – Crane, G. – Majidi, S. (2016), *Part of Speech Tagging for Ancient Greek*, "Open Linguistics" 2, 393–9.

Horrocks, G. (2010), *Greek. A History of the Language and Its Speakers. Second Edition*, Malden – Oxford – Chichester.

Korkiakangas, T. (2016), *Subject Case in the Latin of Tuscan Charters of the 8th and the 9th Centuries.* Helsinki.

Kübler, S. – Zinmeister, H. (2015), *Corpus Linguistics and Linguistically Annotated Corpora*, London – New York.

Reggiani, N. (2017), *Digital Papyrology I. Methods, Tools and Trends,* Berlin – Boston.

Sosin, J. (2010), *Digital Papyrology*, URL: http://www.stoa.org/archives/1263.

Vierros, M. – Henriksson, E. (2017), *Preprocessing Greek Papyri for Linguistic Annotation*, in *Journal of Data Mining and Digital Humanities. Special Issue on Computer-Aided Processing of Intertextuality in Ancient Languages*, ed. by M. Büchler and L. Mellerin, URL: http://jdmdh.episciences.org/paper/view/id/1385. (Version 1 was published in 2016)

Wagner, E.-M. – Outhwaite, B. – Beinhoff, B. (2013), eds., *Scribes as Agents of Language Change,* Boston.

Wynne, M. (2005), ed., *Developing Linguistic Corpora: a Guide to Good Practice*, Oxford, URL: http://ota.ox.ac.uk/documents/creating/dlc.

Joanne Vera Stolk

Encoding Linguistic Variation in Greek Documentary Papyri

The Past, Present and Future of Editorial Regularization

1 Introduction

Linguistic variation in documentary papyri has been noticed by editors since the early days of Papyrology. Some editors make occasional comments about variant spellings[1], others decide not to mention them at all. Kenyon explains his reasons for refraining from marking variation in the introduction to P.Lond. I:

> It is not to be supposed that any human transcript can be entirely free from errors; but the palpable blunders in spelling and grammar with which the papyri abound may be credited in the first instance to the original scribes. It has not been thought worth while to disfigure the pages by appending the warning *sic* to each such violation of conventional rules[2].

In BGU I (1892–1895), the first "truly papyrological edition" according to Van Minnen,[3] the editors added to some transcribed words, such as βιβλείδιον, a note in the critical apparatus saying "*l.* βιβλίδιον" (BGU I 2, n. to l. 17).[4] The method of the 'Berlin editors' is followed by Grenfell and Hunt in their editions published in P.Grenf. II (see p. xii) and P.Oxy. I. They also briefly explain where they consider such a note to be required:

> Faults of orthography are corrected in the critical notes wherever they seemed likely to cause any difficulty.[5]

---

1 E.g. Mahaffy in P.Petr. I 12 (1891), n. to l. 15

2 P.Lond. I (1893), p. vi.

3 Van Minnen 1993, 5–7.

4 The addition of *sic* to unconventional language, as referred to in P.Lond. I (see quote above), is also found in the early BGU editions, next to the regularizations in the apparatus. For example, in BGU II 451 we find τάχειον, *l.* τάχιον (l. 11), ἀσπάσεσθαι with *sic* above ε (l. 9) and ἀσπα|σόμεθά σε with *sic* above σε (ll. 11–12). This is a good example of the challenges faced during digitization of these older editions. All three were initially entered into the DDbDP as regularizations in the apparatus (*l.* τάχιον, *l.* ἀσπάσασθαι and *l.* σοι, respectively). The accusative case σε, however, is normal for the addressee of the verb ἀσπάζομαι and does not require regularization to a dative case, even though that seems to have been suggested by the *sic* in the *ed.pr.*

5 P.Oxy. I (1898), p. xvi.

In 1931, this by then customary practice of regularization was included in the 'Leiden conventions' during the International Congress of Orientalists in Leiden (7–12 September 1931). One would expect that the decision about a unified system of critical signs would be followed by a discussion on how to use them. Whereas several scholars have indeed commented upon the precise meaning and use of some of the signs, such as the underdot, little explanation has been provided about the practice to regularize the Greek language in papyrus documents.[6] Herbert Youtie describes the process as follows:

> Immediately after the text the papyrologist puts a critical apparatus in which he gives conventional equivalents for vulgar or mistaken spellings.[7]

This leaves the most important questions unaddressed, such as 'to which forms should one apply this procedure?' and 'what is a conventional equivalent?'

Regularization implies a norm from which the attested variant deviates. This norm is generally not explicitly formulated in editions and rarely discussed in secondary literature. This makes one wonder whether editors always use the same norms. Whereas the early papyrus editions had to cope with readers that were unfamiliar with the *Koine* Greek language, advances in Greek linguistics and the large corpus of papyrus editions published to date have made most modern readers more accustomed to the features of *Koine* Greek. May this have changed editorial practices? The digitization of papyrus editions in the *Duke Databank of Documentary Papyri* (DDbDP) required a level of standardization across all editions. How did the digitization process influence the consistency of traditional methods? These editorial practices have not been studied before, while they form the basis for our modern tools and digital editions. In order to develop new tools and new methods for digital editing, I consider it important to examine how the current ones are functioning and how we can use existing methods to improve digital technology.

In this paper I analyse the results of a system of editorial regularization which has been in practice for 125 years. The study of editorial practices in the past and present is executed by means of the new *Trismegistos Text Irregularities* tool. This tool collects all editorial interventions that are annotated in the *Papyrological Navigator* (http://www.papyri.info) and allows for detailed searches and analyses of the attestations.[8] I will first give a short overview of the parts of the Leiden conventions that are relevant for the regularization of language and their current application in the digital editions in the *Papyrological Navigator* (section 2). Then, I will discuss the past

---

6 See some notes on the use of critical signs in Hunt 1932 and Youtie 1966. Usually, nothing more is said about the practice of regularization than "give the standard spelling in the apparatus", cf. Schubert 2009, 202.

7 Youtie 1963, 22.

8 For more information about this tool see Depauw – Stolk 2015.

and current use of critical signs and regularizations in the critical apparatus in the original and digital editions (section 3). The possibilities for categorization of variation and different standards are examined in section 4, followed by a concluding section on how we may be able to combine the traditional and modern aims in the development of new digital tools (section 5).

2 The 'Leiden system'

At the 18th International Congress of Orientalists in Leiden (7–12 September 1931), the participants of the Papyrology section discussed the usage of critical signs in editions of inscriptions, papyri and literary authors. They decided on a unified set of conventions, later referred to as the 'Leiden system'.[9] As this was designed to be a universal system for editions of documentary and literary texts, it contained several elements which might seem redundant for editing documentary papyri. Two sets of brackets were chosen to represent scribal omissions and additions to the text, namely the angular brackets ⟨...⟩ for "lacunes" and "additions (lacunes comblées)" and the braces {...} for "interpolations". Of course, interpolations that found their way into the original text through copied manuscripts are not commonly encountered in documentary material. Consequently, these two sets of brackets are in papyrological practice reinterpreted to represent straightforward editorial 'additions' and 'deletions' of letters and words that were forgotten or added superfluously by the scribe of the document for various reasons. The remaining two categories of editorial intervention are "corruptions" and "corrections". Both are indicated in the critical apparatus of documentary texts and are not distinguished formally in papyrus editions. Van Groningen added explicitly that corrections should never replace the text of the papyrus in the transcription (as done with literary texts).[10]

The different types of editorial interventions are all represented in the EpiDoc schema used for marking up textual features in digital editions of inscriptions and papyri.[11] Accordingly, the papyrological conventions used in the *Duke Databank of Documentary Papyri* include the angular brackets for "Characters erroneously omitted by the scribe, added by modern editor", the braces for "Superfluous letters removed by the editor" as well as the option to put regularizations in the critical apparatus.[12] The regularizations in the apparatus can be tagged in different ways in EpiDoc, namely as "Correction of erroneous characters" with the two alternatives marked by `<corr>` and `<sic>` and as "Regularization of dialect or late spellings,

---

9 See *Essai d'unification des méthodes employées dans les éditions de papyrus*, CE 7 (1932), 285–7.
10 Van Groningen 1932, 268.
11 EpiDoc is a TEI-based XML encoding standard developed for digital editions, see Bodard 2010.
12 http://papyri.info/conventions.html, accessed on 22 May 2017.

etc." marked by `<orig>` and `<reg>`.[13] Both are used in the collaborative online editing environment of the *Papyrological Navigator*, called the *Papyrological Editor*.[14] This platform uses a non-XML representation of the EpiDoc schema, called 'Leiden+', in order to facilitate easy entry of new texts by its users.[15]

All editorial conventions used in Leiden+ are explained to the user in a set of online guidelines.[16] The Leiden+ Documentation tells the digital editor to distinguish between a "spelling correction" to be used for "correction of outright scribal error"[17] and an "orthographic regularization" to be used for a "non-standard orthographic form".[18] According to the guidelines, critical signs should be used for spelling corrections as well, which reduces the practical difference between the four categories into two basic types. The PN is thus expected to encode

1. 'corrections' by means of critical signs (for additions and omissions) and in the apparatus (for substitutions and more complex cases), and
2. 'regularizations' of non-standard forms in the apparatus.

3 Editorial regularization in practice

Although papyrologists have agreed on the methods to be used in papyrus editions, as described above, the application of these basic principles is not self-evident. Herbert Youtie already stated in his prolegomena to the textual criticism of documentary papyri:

> it is a far cry from subjective opinion to objective reality, although no hint of this difficulty is ever betrayed in the definition of the signs that we find in papyrological manuals.[19]

---

13 For more information about these two and other possible editorial interventions see http://www.stoa.org/epidoc/gl/latest/app-alltrans.html, accessed on 22 May 2017.

14 http://papyri.info/editor.

15 Baumann 2013, 102–4; Sosin 2010.

16 The Leiden+ guidelines (http://papyri.info/docs/leiden_plus) have been subject to revision since the start of the editorial interface to the *Papyrological Navigator*. The unfortunate decision to display the corrected reading in the text and the original in the apparatus has been changed to the common practice in editions to show the original text in the transcription and regularizations in the apparatus. However, this technical change still has some consequences for the display of critical signs, line breaks and accents of regularized words that were entered before the change. Some attempts have been made to clarify the distinction between corrections and regularizations in the guidelines with varying results, cf. section 3.

17 http://papyri.info/docs/leiden_plus#spelling-correction, accessed on 22 May 2017.

18 http://papyri.info/docs/leiden_plus#orthographic-regularization, accessed on 22 May 2017

19 Youtie 1974, 64.

While this may apply to all critical signs, it is especially true for the editorial regularizations of the language found in papyrus documents. I will illustrate this by some examples mentioned below.[20]

Following the basic distinctions available in EpiDoc (see section 2), I will distinguish between the so-called 'corrections' indicated by means of critical signs and in the apparatus (section 3.1) and 'regularizations' in the apparatus (section 3.2). The starting point for this comparison is the database of TM *Text Irregularities*, which contains a collection of all editorial regularizations in papyrus editions in the PN.[21] There are two stages to take into account: the regularization indicated in the *editio princeps* and the annotation in the digital edition in the PN. This method will allow me only to quantify the outcomes of the second stage of this process. It should be noted that the digital edition in the PN is not always a true replica of the original edition, as more regularizations have been added in an attempt to level out the differences in conventions between various (older) editions. Hence, for every example mentioned below, I will also compare the digital regularization with the one in the original edition in order to reflect on possible differences between the two stages of editing.

3.1 Corrections and critical signs

The EpiDoc schema offers the possibility to distinguish between corrections of scribal errors and orthographic regularizations (see section 2). The application of a special 'correction' tag results in the addition of (corr) after the corrected form in the apparatus of the digital edition. In practice, it has never been in frequent use and some earlier instances have been automatically converted into regularizations. The remaining 140 corrections might have slipped through the net at an earlier stage or may have been added later, as users are still confronted with guidelines mentioning this option.[22]

A closer look at the instances that are encoded as correction at the moment reveals that a significant part of them does not seem to fit the definition of "outright scribal error". Regularizations of interchanges resulting from phonological mergers, such as ἰς to εἰς in O.Claud. IV 723 and Παραδίσου to Παραδείσου, λεῖβα to λίβα and [ἀπ]οδόσω to [ἀπ]οδώσω in SB XXVI 16796,10–11, 16, are regularly found among these

---

corrections (50 times).[23] Morphological regularizations are also common (41 times). For some of those, it is possible to see why the (digital) editor regarded them as scribal errors. For example, in BGU XVII 2682,6, 9–10, the scribe mechanically added the standard accusative object χωρίον ἀμπελικόν, whereas in this particular construction ([ὁ]μολογῶ ... μερίδαν | μίαν χωρίον ἀμπελικόν) the noun phrase should have been a genitive partitive to the object μερίδαν | μίαν.[24] In P.Gen. IV 192,10–11, the pronoun σοι was inserted too early and ended up with the wrong verb: ὁμολογῶ ἔχειν σοι καὶ | χρεωστεῖν instead of ὁμολογῶ ἔχειν καὶ | χρεωστεῖν σοι. Printed editions do not make a distinction between mechanical scribal errors and other regularizations, although they sometimes provide an explanation for the variation in the commentary (as was done for BGU XVII 2682, n. to l. 10). Apart from those occasional comments, the interpretation of the distinction between regularization and scribal error depends largely on the person digitizing the edition. The phrase σὺν ναύλαις κὲ ἐκαταστῆς was regularized as "*l.* ναύλοις καὶ ἑκατοσταῖς" in the apparatus of P.Jena II 8,7, but ναύλοις was entered into the PN as a correction, καὶ as regularization and ἐκατοστῆς as regularization (probably mistakenly for ἑκατοσταῖς). Obviously, the distinction between the two types of regularizations creates a great challenge for the digital editor, especially without a clear definition of 'scribal error' at hand.

Besides the special correction tag, simple scribal errors can also be indicated with critical signs according to the guidelines (see section 2).[25] The angular brackets (for editorial additions) and braces (for editorial deletions) are in common use in both printed and digital editions. In TM *Text Irregularities*, we collected a total of 6,920 instances of the use of angular brackets and 3,063 attestations of braces in the digital editions in the PN. Both of them are primarily used for scribal omissions and additions of whole words, amounting to 66% and 80% of the instances of the angular brackets and braces respectively. This also forms the main distinction between the use of critical signs in the text and regularizations in the apparatus: the critical signs mark additions and deletions of whole words, while regularizations are almost exclusively limited to parts of words.[26] However, the critical signs are also used for single letters

---

23 Based on the collection in TM *Text Irregularities* I made a list of the 'corrections' that are more likely to be the result of phonological changes in the language (cf. 4.1), so that these could be converted into regularizations in PN. Josh Sosin replied to me that these corrections will be converted, but the option to distinguish between different types of errors is going to be maintained in the PE in the future (personal communication, 7 June 2017).

24 See also Vierros 2012; Stolk 2015, 268–71.

25 http://papyri.info/docs/leiden_plus#leiden-angle-brackets;
http://papyri.info/docs/leiden_plus#leiden-braces;
http://papyri.info/docs/leiden_plus#spelling-correction, accessed on 23 May 2017.

26 If regularizations in the apparatus are used for the addition of several words, the angular brackets are sometimes added to the apparatus entry as well, e.g. χειρογραφεῖσα, *l.* χειρογραφ⟨ία ἁπλῆ γραφ⟩εῖσα in P.Oxy. XXXIV 2724,20.

and parts of words and in this usage they often overlap with the regularizations. Around 82% of the angular brackets and braces put around part of a word are in fact used to indicate interchanges at a phonological and/or morphological level, such as ⟨ε⟩ι or {ε}ι (160 times) and the addition and omission of final -ς (237 times) and -ν (198 times). If one would want to achieve a meaningful difference between the use of critical signs and regularizations in the apparatus, critical signs in the text should not be used for orthographic and morphological interchanges affecting only a single letter or part of a word.[27]

3.2 Regularizations in the apparatus

Regularizations are traditionally indicated with '*l.*' for *lege* "read" in the apparatus of an edition. They make up the majority of all instances of editorial linguistic intervention in papyrus documents (92 %), amounting to more than 120,000 instances in all digitized papyri. Most of the editorial regularizations concern orthographic variation caused by changes in the pronunciation of *Koine* Greek (70%). Another significant part of the regularizations affects the spelling and use of morphemes (26%), such as case and verb endings. I divide the variation at a morphological level into two types:

1. morphological interchange between different declensions or conjugations, such as the variation between an accusative singular in -α and -αν for consonants stems or between the sigmatic and root aorist inflection of certain verbs,[28] and
2. morphosyntactic variation between the use of morphemes in a particular syntactic context, such as between a genitive or a dative case to express the recipient of a verb of giving or between an indicative or subjunctive following the conjunction ἵνα.

Both types occur among the regularizations in the apparatus. In some cases, morphological or morphosyntactic variation may be related to phonological merger as well. An example of this is the frequent interchange of ο and ω, of which one third of the instances are found in case endings (e.g. τόν / τῶν) and two thirds in other positions (e.g. ὠκτώ / ὀκτώ). It is, therefore, not always easy to distinguish different types of variation based on the level of language organization that they apply to.

Almost 40% of the regularizations of orthographic variation concern the interchange of ι and ει. For most of these variant spellings, regularization is not strictly necessary in order to understand the meaning of the word. Still, there are many forms

---

27 Apart from the large group of common phonological and morphological irregularities, the remaining 20% may concern a relatively high portion of potential 'scribal errors'. The problematic identification of these 'scribal errors' will be addressed in section 4.1.

28 GIGNAC 1981, 45–6 and 290–7.

for which one spelling has been regularized consistently in (almost) all instances throughout the corpus, such as ἴκοσι to εἴκοσι, ἔχις to ἔχεις, ἐλθῖν to ἐλθεῖν etc. This is partly due to the addition of regularizations during the digitization process. For example, in the *ed.pr.* of P.Oxy. XLIII 3117 interchanges between ι and ει are only regularized when they could be confusing (e.g. ἐπί to *l.* ἐπεί in ll. 6 and 14), but many others have been added in the digital edition, such as to βιβλεία in l. 4, κοινωνῖν in l. 5 and ἀποκρείνασθαι in l. 6. The few instances where regularization in the PN is lacking may be caused by human error, such as the typo 'πάλειν for πάλειν' rather than πάλιν in the digital edition of in P.Oxy. XLIII 3117,13–14 (http://papyri.info/ddbdp/p.oxy;43;3117), or the regularization of περεί to περί in the *ed.pr.* of SB XX 14990,15,[29] which seems to have been overlooked in the *Sammelbuch* and the digital edition.

There are also words for which the standard spelling may be more difficult to establish. According to classical rules, the suffix of the derived noun ὑπερφύεια "excellency" is spelled with ει.[30] This is also the spelling which is found in the majority of the VI- and VII-century papyri, such as the attestations in P.Oxy. I 135–138 published in 1898. The alternative spelling ὑπερφύια is not regularized in P.Oxy. I 144,4, nor in the five papyri P.Cair.Masp. I 67003, 67005–67008, published in 1911. Regularizations of ὑπερφύια do occur in editions that were published later, such as P.Ross.Georg. V 34,2 (published in 1935), CPR XXIV 27,17 (published in 2002), and P.Oxy. LXX 4790,16, 19 and 30 (published in 2006). The alternative spelling in P.Oxy. I 144,4 became eventually regularized in the online edition. P.Cair.Masp. I 67003, 67005–67008 remain without regularization in their online editions.[31] Remarkably, a regularization of the common form ὑπερφύεια to ὑπερφύια was also added to the digital edition of P.Lond. III 774–778.[32] Whereas the earlier editions seem rather modest with regularizations of words that can be perfectly understood without, the growing need for consistency may have extended regularization to be applied to all 'non-standard' forms without agreement on the definition of 'non-standard'.

The Leiden conventions were designed to do reduce variation in editorial practices. The common format of a transcription with a critical apparatus containing regularizations becomes indeed the standard for all editions, but the variation in regularization practices continues in printed editions after 1931. The word βιβλιοφυλάκιον "archive" is spelled as such in 22 papyri and as βιβλιοφυλάκειον in six papyri between the II and IV centuries AD. The spelling βιβλιοφυλάκειον is regularized to βιβλιοφυλάκιον in the edition of P.Diog. 20, 6, and the online editions of SB VI 9625,23,

---

29 Herring 1989, 31–3.

30 Cf. Palmer 1945, 54.

31 Perhaps accidentally; or because the alternative spelling seems to have been the norm in the Dioscorus archive.

32 These documents originate from the Apion archive, just as most of the other documents with the word ὑπερφύεια, and they show the spelling that is normally found in this archive. It is, therefore, not clear what the regularization was based on.

PSI V 454,19, and P.Tebt. II 318,23; the normalized spelling is also found in the index of the last two editions. For the two editions that remain without regularization (P.Gen. I² 144,23; P.Hamb. I 16,22), the spelling βιβλιοφυλακεῖον was used both in the texts and indices of the original editions. Strikingly, the more common spelling βιβλιοφυλάκιον was even regularized to βιβλιοφυλάκειον in the first editions of P.Oxy. XXXIII 2665,17 and 19, P.Fam.Tebt. 15,iii,79, and P.Hamb. IV 244,12, and there are also editors that supplement the word in this spelling in abbreviations or lacunae (see BGU III, p. 2 to BGU I 243,15, taken over in Chr.M. 216; Chr.M. 217,9, and P.Fam.Tebt. 29,44).

Inconsistent regularizations, such as the ones mentioned above, often require careful analysis to determine whether this apparent lack of consistency can be justified in any way in each of the given situations and based on the material that the editors had at their disposal. Similarly, complicated situations arise when one attempts to regularize morphosyntactic variation. The phrase ἐάν σου τῇ τύχηι δόξῃ "if it seems right to your fortune" occurs regularly in petitions from the II and III centuries AD. The second person singular pronoun is usually in the genitive case (σου), but it is also attested in the dative case (σοι). The dative σοι is regularized into a genitive σου in SB XXIV 15915,6, while SB XVIII 13732,13, regularizes the common genitive into the dative in this phrase.[33] Confusion about the use of the dative or genitive case in these types of constructions is common among both scribes and editors and regularization is often far from straightforward.[34]

Inconsistent regularizations are usually caused by a lack of agreement about the method of standardization. Differences between older editions have not always been levelled out during the digitization process and they might even have gotten worse in some of the more complicated examples mentioned above. Some editions take a more extreme approach than others when it comes to choosing a method for regularization. Common itacistic spellings, such as εἴνα and ἱς, are often regularized in papyrus editions, but not in the editions of the Mons Claudianus *ostraka*. This is probably because these particular interchanges are very common in these *ostraka* and regularization seems unnecessary.[35] This practice is not entirely consistent throughout the volumes (e.g. O.Claud. IV 723 and 839 regularize ἱς, but O.Claud. IV 724 and 840 do not). Regularizations have been added during the digitization process in accordance with other papyrus editions, but the end result is still far from uniform (e.g. ἱς has been regularized in the digital editions of O.Claud. II 248 and 276, but not in O.Claud. II 363 and 383).

Comparison between texts in the same volume and among other parallel texts is a common practice, but it is not the main method of regularization in most papyrus editions. The word νοσοκομεῖον "hospital" is attested in full in 14 papyri dated to the VI and VII centuries. Only one of those attestations is spelled with ει (SB I 4668,4), as

---

33 See STOLK 2017, 196–7 with n. 31.

34 For more examples see STOLK 2015 and 2017.

35 Compare the comment by Grenfell and Hunt in P.Oxy. I, cited above in section 1.

is also common in modern Greek, while all the others write νοσοκομῖον.[36] Based on comparison to contemporary documents, no regularization seems required for the other instances. In reality, regularizations to the standard spelling νοσοκομεῖον are found in original editions (e.g. P.Bodl. I 47,12, 20 and 26; CPR XXII 2,1, 5 and 9) and digital editions (e.g. P.Amh. II 154,2 and 8; P.Lond. III 1324,7), while others remain without any form of regularization (e.g. P.Oxy. XVI 1898,19 and 38; P.Oxy. XIX 2238,18).[37] This combination of different methods will inevitably lead to more inconsistencies within and between printed and digital editions in the future.

4 Standardization

Past and current approaches have not resulted in a clear distinction between 'scribal error' and 'non-standard variant' in the PN (see 3.1). The question remains whether it is possible to distinguish scribal errors from other types of variation and whether we should want to make such a formal distinction in (digital) editions (4.1).

It has been shown that regularization of variation due to phonological, morphological and morphosyntactic changes is not always consistent (see 3.2). Editors may use different methods to identify the norm and, consequently, these norms may differ from each other. In section 4.2, I will discuss various possibilities for establishing a standard for comparison.

4.1 Scribal errors

The traditional aim of textual criticism is the "Herstellung eines dem Autograph (Original) möglichst nahekommenden Textes".[38] Any corruptions to the text are caused by "the inability of scribes to make an accurate copy of the text that lay before them".[39] Hence, any form of scribal intervention can be regarded as a mistake.[40] Similar phenomena, such as misreading of the exemplar, orthographic variations and accidental alterations, occur in duplicate papyri, but not all documentary papyri are the result

---

36 The spelling νοσοκομῖον is also common in Coptic, cf. Förster 2002, 549, and see e.g. CPR IV 198,16 and 21.

37 Supplements for abbreviations show the same variation. The spelling with ι is supplemented in abbreviations in Stud.Pal. III 314,1, Stud.Pal. VIII 791,1, and 875,2, while ει has even been supplemented in papyri where the spelling with ι is found elsewhere in the same text, see CPR XXII 2,1, 5, 9 and 11; P.Oxy. LXI 4131,16 and 39.

38 Maas 1950, 5.

39 Reynolds – Wilson 1991, 222.

40 Some examples of such (deliberate or accidental) mistakes are given in the list in Reynolds – Wilson 1991, 222–33.

of copying.[41] Therefore, our definition of scribal error has to be different from the one used for copying literary texts.

Papyrus documents are the product of their own time and not the result of several centuries of transmission. Therefore, changes in the language do not need be regarded as scribal or copying errors in documents. Still, a division between scribal error and linguistic variation is commonly applied in linguistic approaches. Variation in the written language can be used to reconstruct changes in the history of the spoken language. In order to do that, significant variations, i.e. interchanges reflecting the spoken language, have to be separated from "Verschreibungen"[42], "garbage errors"[43] or "manifest blunders"[44]. Gignac identifies this difference between "phonetically significant variation" and "sheer mistakes and slips of the pen" by the principles of frequency and regularity:

> If certain letters or groups of letters interchange only rarely and irregularly, there might be another explanation.[45]

His other explanations include (a) anticipation and repetition, (b) inversion, (c) mechanical reproduction, (d) analogical formation and (e) etymological analysis.[46] These examples of variation which occur irregularly and do not seem to reflect the spoken language can be described as 'scribal errors'. Scribal errors of this type can usually be explained by common cognitive processes.[47]

Mechanical and cognitive processes may explain the appearance of scribal errors, but they do not constitute a comprehensive categorization or definition of the phenomenon itself. Haplography and dittography, for instance, are prime examples of the cognitive processes of anticipation and repetition (a). However, the simplification and gemination of consonants can also be explained by "the identification in speech of single and double consonants".[48] Hence, the example of "outright scribal error, e.g. στ[ρ]αττεός for στρατηγός" given in the Leiden+ documentation[49] can also be explained by hypercorrective gemination of the consonant, the phonetic similarity of ε and η and the omission of γ in the pronunciation as glide.[50] Even the loss of a full

---

41 For a typology of scribal errors in duplicate papyri see Yuen-Collingridge – Choat 2010.

42 Kapsomenakis 1938, 4.

43 Lass 1997, 62.

44 Jannaris 1907, 68.

45 Gignac 1976, 57 and 59.

46 Gignac 1976, 59.

47 Cf. Kapsomenakis 1938, 4.

48 Gignac 1976, 154–5.

49 http://papyri.info/docs/leiden_plus#spelling-correction, accessed 22 May 2017. A better example is `<:τιμὴν|corr|τμμὴν:>`.

50 Gignac 1976, 242–7 and 71–5.

syllable may sometimes have a phonetic explanation.[51] Inversion (b) is another problematic category. Although the transposition of two letters may result "in spellings like *atmoshpere* which do not reflect an actual spoken form", metathesis of a vowel and resonant, especially ρ, is relatively frequent in the papyri and may have had a parallel in speech.[52] Mechanical reproduction (c) seems to identify a type of variation that is indeed limited to the written language, but the two remaining categories on Gignac's list are not scribal errors strictly speaking either. Analogical formation (d) may not be caused by phonological changes, but can be indicative of morphological change in the spoken language, as is also acknowledged by Gignac.[53] When a form can be explained by changes in the spoken language (phonological or morphological), it should not be classified as a scribal error according to the definitions mentioned above. Etymological analysis (e), such as the spelling of ἐκ- in compounds before a voiced consonant, may not be relevant for the actual pronunciation of the word in later periods, but this change in orthographic conventions is better classified as orthographic variation than as a mechanical scribal error.

Mechanical scribal errors in papyrus documents have received little study in their own right. Negative definitions prevail in the secondary literature aiming at the reconstruction of the original text or the spoken language. Gignac gives an excellent introduction to his method, but his overview of orthographic variations that are not phonetically significant cannot be used as a typology of scribal errors in documentary papyri.[54] Editors should feel free to discuss causes for variation in their commentaries and digital editors might want to continue experimenting with these distinctions, but it would be better to treat possible scribal errors in the same way as other types of variation in order to secure stable future reference to all variant forms.

4.2 Different standards

Regularization implies the use of a standard. Every editor who regularizes the language found on a papyrus compares the attested words and constructions with a certain norm. How and why this norm is chosen is usually not stated explicitly, but can be inferred to a certain extent from the patterns of regularization observed above (see 3.2). There seem to be two main sources for comparison:
1. external sources, such as rules described in dictionaries, grammars and text books, and

---

51 Cf. GIGNAC 1976, 312–3.
52 GIGNAC 1976, 59 and cf. pp. 314–5.
53 GIGNAC 1976, 59.
54 GIGNAC 1976, 57–60.

2. internal sources, such as other instances in the text itself or parallel texts that are ideally closely related in contents and context.

Regularization to νοσοκομεῖον, for example, was probably based on external criteria in most instances, since this spelling is rarely found in contemporary papyri. The spelling of ἱς, on the other hand, may have been left without regularization in the *ostraka* from Mons Claudianus based on comparison to other *ostraka* from the same area. As long as the attestations found in close parallels corroborate the external standards, editorial regularization of variant spellings tends to be consistent. As soon as both variants seem to be in regular use in contemporary papyri, such as with βιβλιοφυλάκ(ε)ιον and ὑπερφύ(ε)ια, different editorial principles and methods may lead to conflicting results.

It is not true that classical norms were especially used in the early days of papyrology and comparison with contemporary documents is an entirely new phenomenon. Variation in regularization practices is particularly common in early papyrus editions and classical norms are not consistently applied at all (cf. 3.2). Recent studies of the language of the papyri, often from a variationist perspective, have raised awareness of the possibility that scribal variation could be explained by its context.[55] This may have led some editors to consider more context-sensitive methods, but also more practical considerations may have prevented editors from regularizing spellings that occur very frequently in a specific group of documents. The variationist idea that linguistic variation is dependent on its context is not an entirely new concept to papyrologists. The principle of comparison with parallel texts for understanding and supplementing another papyrus has been in use for a long time. In order to interpret the language used in papyri, Youtie suggests the use of dictionaries, grammars and "an unremitting search for parallels".[56] He further notes that

> U. Wilcken has somewhere characterized papyrology as a "Parallelenjagd". No term could be more apt. A good share of the papyrologist's working time is devoted to searching for parallels.[57]

Parallel examples are essential for a papyrologist to get familiar with the language and contents of different types of documents, to date the text and to identify the standard clauses used at different times and places.[58] Even though this method has been used for many years to interpret new texts and to supplement words and phrases

---

55 See the papers in EVANS – OBBINK 2010; LEIWO – HALLA-AHO – VIERROS 2012; CROMWELL – GROSSMAN forthcoming.

56 YOUTIE 1974, 33–7.

57 YOUTIE 1974, 42 n. 39.

58 Cf. TURNER 1980, 59–61. The 'hunt for parallels' is one of the main incentives for the digitization of papyrus editions, because it makes it easier for papyrologists to search for parallels in a large corpus of published papyri.

in fragmentarily preserved papyri, it is not always deemed suitable as a standard for linguistic comparison. Classical orthography and morphology are often understood to be the only proper standard for the language used in regularizations, in supplements of abbreviations and in lacunae.[59] Kapsomenakis already voiced his concerns about the artificial norms that tended to be applied to the Greek language in papyri:

> Übrigens hat eine volksmäßig frei entwickelte Sprache ihre eigenen Gesetze, denen sie folgen muß, wenn sie ihre Aufgabe, der praktischen Verständigung zu dienen, erfüllen will. Die Vulgarismen dürfen also diesen Gesetzen nicht widersprechen. Weiter hat die Verkennung der Rechte der Volkssprache dazu geführt, daß man viele Schreiberfehler entdeckte, die man nach der Methode beseitigen zu müssen glaubte.[60]

Classical Attic norms continued to be used as the standard for orthography and morphology in post-classical periods, but it seems difficult to justify applying anachronistic norms in cases in which a variant form is frequently or even normally used in *Koine* Greek. Lack of the awareness of the norms for the language used in papyri can easily lead to misplaced regularizations, reconstructions and even readings.[61] The discrepancy between classical Attic and contemporary usage as the norm for editorial regularization is probably caused by a general lack of information about contemporary norms, as has also been pointed out by Youtie:

> But it is perhaps lack of linguistic information which trips us most often. Sometimes this takes the form of insufficient regard for the general laws of Hellenistic Greek, sometimes it is simply failure to search out the similar passages which are available in other papyrus texts. Whatever its cause, it has a crippling action capable of twisting our texts into fantastic shapes.[62]

Knowledge about *Koine* Greek in general and the linguistic norms applied in papyri in particular are essential ingredients for a good papyrus edition and may help to prevent many reading errors and problematic restorations. On the other hand, the standards for orthography, morphology and morphosyntax in *Koine* Greek have still received little attention in research to date and there is no reference work that editors can use to identify a standard for every word or construction. These norms can, therefore, only be identified by manual comparison among a selection of documents. This creates the typical gap between the use of external sources based on classical Greek and the contemporary internal evidence.

---

59 Linguistic inconsistencies in the practices of restoration of the text in lacunae are clearly pointed out in Evans forthcoming. I thank Trevor Evans for kindly sharing this unpublished paper with me and for sharing his thoughts about these issues.

60 Kapsomenakis 1938, 4.

61 The problematic consequences of the practice to restore (and even read) classical Greek forms where they have not been written originally are illustrated in Clarysse 2008 and Youtie 1974, 8–10 and 13–16.

62 Youtie 1974, 13.

Koine Greek has never become a general standard for editorial regularization, although there are some exceptions. An example of such a well-known orthographic norm in *Koine* Greek is the spelling of the verbs γί(γ)νομαι 'to be, to become' and γι(γ)νώσκω 'to know'. Mayser and Schmoll state that the spellings γίνομαι and γινώσκω are used without exception in the Ptolemaic papyri and Gignac confirms that these are also the normal spellings in the Roman period.[63] Accordingly, the spelling γίνομαι is usually not regularized and the *Koine* Greek spelling is used in most supplements of abbreviations of the verb.[64] Still, regularization to γίγνομαι is found in about a dozen editions (e.g. P.Bodl. I 17,i,9; P.Haun. II 22,5; P.Oxy. LXIV 4441,x,27; P.Petra I 4,5) and has occasionally been added to digital editions as well (e.g. O.Claud. IV 798,6; P.Stras. VIII 772,6, 9, 15 and 21). In contrast to the relatively limited number of regularizations of γίνομαι, the verb γινώσκω has been regularized to γιγνώσκω in more than a hundred instances. Most of these regularizations, however, concern verbs with other spelling irregularities (almost 90%), such as γεινώσκιν to γιγνώσκειν (e.g. P.Col. X 278,4; SB XXIV 16290,2 and 16291,4).[65] When regularizing these other aspects, the idea of the classical standard seems to have overruled *Koine* Greek spelling conventions. The verb γίνομαι is also frequently spelled as γείνομαι, but this rarely provoked regularization to the classical spelling of the consonants. The fact that the spelling of the verb γίνομαι often serves as the prime example of language change in *Koine* Greek, may have convinced editors to take the *Koine* Greek spelling as the standard for this verb more often.[66] The differences in regularization between these two comparable verbs clearly illustrate the competing principles of regularization.

5 Towards a new approach

In the previous sections, I have illustrated the various practices and principles for editorial regularization as they have been used up till today. Editorial regularizations

---

63 Mayser – Schmoll 1970, 15 and 156; Gignac 1976, 176; see also LSJ s.v.

64 Supplements of abbreviations and lacunae are other sources for editorial disagreement on linguistic variation. Different principles, such as regularization to classical orthography and comparison within the document or to other contemporary documents, are used by different editors. Since there is no current method to search for attestations in the real text only, search results are often biased for standard forms found in supplements and the apparatus.

65 Paul Schubert regularized γεινώσκιν to γινώσκειν in the *ed.pr.* of SB XXIV 16290 and did not put a regularization to γεινώσκι[ν in the *ed.pr.* of SB XIV 16291, see Schubert 1997, 193–4. Clearly, the need for standardization of regularization practices is not only felt during the digitization process, but also in large collections of papyrus editions such as the *Sammelbuch*.

66 The *sic* of the editors behind the unusual spelling and morphology τὰ γιγνώμενοι in O.Edfou II 318,7, was even regularized to the *Koine* Greek spelling *l.* γινόμενα in the digital edition.

in the apparatus are used to indicate phonological, morphological and morphosyntactic variation (3.2), while critical signs, such as angular brackets and braces, are mainly used by the editors to mark the addition and omission of one or more words (3.1). When brackets and braces are applied to single letters or parts of words, their function largely overlaps with the regularizations in the apparatus. More study is needed to separate accidental scribal errors from other types of variation in the papyri (4.1). The same applies to establishing contemporary standards for *Koine* Greek (4.2). Both goals are worthwhile pursuing in separate studies in order to gain a better understanding of the use of language in papyri, but such a distinction between different types of variation or different standards might not be essential for establishing a more consistent practice of encoding linguistic variation.

The question comes down to what we would like to achieve with editorial regularization in papyrus editions. Are we trying to correct accidental scribal mistakes in the way the scribe would have wanted to? Are we normalizing the language to conservative or contemporary standards? Or are we just helping the classically schooled modern reader to understand a text written in a different variety of Greek? This last idea was probably an important reason to start providing standard Attic equivalents in the apparatus, as Turner explains:

> The critical apparatus [...] can also usefully show how the editor understands his text. The word 'read' or symbol *l.* = '*lege*' need not mean that the Greek is incorrect: it is a sign of how it can be interpreted in terms of standard Attic Greek.[67]

The fact that Turner has to explain what is not meant by this sign immediately points out that the use of the word "*lege*" can be misleading. The command "read" is easily interpreted as a correction rather than an equivalent. This inherent ambiguity is worth noting here. Other, more appropriate, signs should be considered for future printed editions. For digital purposes, however, it would be better to take a different approach altogether. As the apparatus shows 'how the editor understands his text', the ideas about what should be explained in the apparatus and what not can differ significantly from one editor to the other. Some editors may follow this practice very strictly and always provide standard equivalents, whereas others may think that this is only necessary for forms that are less common and more difficult to understand for the reader of the edition, such as in the earlier editions of the Oxyrhynchus papyri. This results in a pragmatic and fluid norm for encoding variation.

Fluid norms are not ideal in a digital environment. That is why it was attempted to make regularization more consistent in the DDbDP and in the *Papyrological Navigator*. Modern editors and the methods designed for the *Papyrological Editor* have succeeded in standardization of editorial practices in digital editions to a large extent, but consistent regularization is not always a straightforward procedure, as I have

---

67 TURNER 1980, 71.

shown in the sections 3.2 and 4.2. Variation may be governed by various factors and this means that the chosen method for regularization may sometimes determine the outcome. This causes problems, because there are no guidelines describing a particular methodology for regularization in papyrus documents.

Digital technology, however, can do more than standardizing the practices of printed editions. I can identify three main aims for encoding linguistic variation in papyrus editions:

1. to show readers of the edition how the editor interprets uncommon forms,
2. to help papyrologists to search for parallels of words and phrases in various spellings, and
3. to provide useful data for linguists studying the *Koine* Greek language.

The current practices in editorial regularization and the search interface of the PN do not fulfil each of those aims equally well.[68] The traditional method of regularization is not suitable to encode variation consistently and objectively. In order to achieve objectivity we need to apply the same treatment to all forms rather than to rely on the judgements of individual editors to identify which forms are 'uncommon' enough. One could do this by providing a reference to a headword, i.e. a lemma, to every word that is attested on a papyrus. It is already possible in EpiDoc to annotate a 'lemma' attribute to every linguistic 'token'.[69] A hyperlink to a lemma can be very helpful for less experienced users and additional morphological annotation would give an opportunity to the editor to explain how every form should be interpreted. The encoding of lexical and/or morphological information for every word can be similar to the creation of onomastic and prosopographical references to every proper name.[70] The lemma could be in classical orthography, as it is not meant as a correction or regularization, but only as a reference point for all variant spellings. A search query would yield all attested variants of the lexeme or morpheme in question. Such an overview of attested variants and their chronological and geographical contexts will show which form might have been in common use at any given time. Full text search queries should ideally separate between the results based on real attestations on a papyrus and results including supplements of abbreviations and restorations in lacunae

---

68 Current search results in *Papyrological Navigator* do not give the number of attestations, but only the number of texts in which one or more attestations can be found. Furthermore, the *Papyrological Navigator* does not allow searches in the main text only; comments and regularizations in the apparatus are always included among the search results. Hence, the number of found attestations of standard forms is biased due to the high number of regularizations in the apparatus, as supplemented abbreviations and as restorations in lacunae. Real attestations can only be distinguished from examples in lacunae and in the apparatus by going through the search results manually. This especially affects the practicalities of the second and third point of the aims mentioned above.

69 See BODARD 2010.

70 Cf. *Trismegistos People*, http://www.trismegistos.org/ref/index.php, and BROUX – DEPAUW 2015.

added by editors. This will provide the papyrologist with a more realistic picture of the language used in papyri and this will benefit new editions in the future.

Marking linguistic variation is not a bad idea in itself, nor is the attempt to standardize editorial practices in digital editions. However, in order to reach the full potential of these approaches, they need to be applied more rigorously and more objectively. Editorial regularizations that have been annotated up to now should not be discarded, but can be used to establish automatic recognition of the lemmata and their potential variants. Once such a digital tool is functioning properly, only the most uncommon forms would still need to be annotated manually, comparable to the original practice of regularization. There are different technological solutions and several possible platforms that would be suitable to achieve these aims. Until that moment, papyrologists and linguists will be able to explore the rich source of linguistic variation available in *Trismegistos Text Irregularities*, a collection of a long history of editorial regularization.

6 Bibliography

Baumann, R. (2013), *The Son of Suda On-Line*, in *The Digital Classicist 2013*, ed. by S. Dunn and S. Mahony, London [BICS Suppl. 122], 91–106.

Bodard, G. (2010), *EpiDoc: Epigraphic Documents in XML for Publication and Interchange*, in *Latin on Stone: Epigraphic Research and Electronic Archives*, ed. by F. Feraudi-Gruénais, Lanham, 101–18.

Broux, Y. – Depauw, M. (2015), *Developing Onomastic Gazetteers and Prosopographies for the Ancient World through Named Entity Recognition and Graph Visualization: Some Examples from Trismegistos People*, in *Social Informatics. SocInfo 2014 International Workshops, GMC and Histinformatics (Barcelona, Spain, November 10, 2014)*, ed. by L.M. Aiello and D. McFarland, Cham, 304–13.

Clarysse, W. (2008), *The Democratisation of Atticism: Θέλω and ἐθέλω in Papyri and Inscriptions*, ZPE 167, 144–8.

Cromwell, J. – Grossman, E. (forthcoming), eds., *Beyond Free Variation: Scribal Repertoires from Old Kingdom to Early Islamic Egypt*, Oxford.

Depauw, M. – Stolk, J. (2015), *Linguistic Variation in Greek Papyri. Towards a New Tool for Quantitative Study*, GRBS 55, 196–220.

Evans, T.V. (forthcoming), *Textual Criticism of Greek Papyri*, in *Oxford Handbook of Greek and Latin Textual Criticism*, ed. by W. de Melo and S. Scullion, Oxford, in press.

Evans, T.V. – Obbink, D.D. (2010), eds., *The Language of the Papyri*, Oxford.

Förster, H. (2002), *Wörterbuch der griechischen Wörter in den koptischen dokumentarischen Texten*, Berlin – New York.

Gignac, F.T. (1976), *A Grammar of the Greek Papyri of the Roman and Byzantine Periods. I: Phonology*, Milano.

Gignac, F.T. (1981), *A Grammar of the Greek Papyri of the Roman and Byzantine Periods. II: Morphology*, Milano.

Herring, D.G. (1989), *New Ptolemaic Documents Relating to the Shipment of Grain: Five Naukleros Receipts and an Order to Sitologoi*, ZPE 76, 27–37.

Hunt, A.S. (1932), *A Note on the Transliteration of Papyri*, CE 7, 272–4.

Jannaris, A.N. (1907), *Latin Influence on Greek Orthography*, CR 21, 67–72.

Kapsomenakis, S.G. (1938), *Voruntersuchungen zu einer Grammatik der Papyri der nachchristlichen Zeit*, München.

Lass, R. (1997), *Historical Linguistics and Language Change*, Cambridge.

Leiwo, M. – Halla-aho, H. – Vierros, M. (2012), eds., *Variation and Change in Greek and Latin*, Helsinki.

Maas, P. (1950), *Textkritik. 2., verbesserte und vermehrte Auflage*, Leipzig.

Mayser, E. – Schmoll, H. (1970), *Grammatik der griechischen Papyri aus der Ptolemäerzeit. Band I. Laut- und Wortlehre. I. Teil: Einleitung und Lautlehre. Zweite Auflage*. Berlin.

Palmer, L.R. (1945), *A Grammar of the Post-Ptolemaic Papyri*, London.

Reynolds, L.D. – Wilson, N.G. (1991), *Scribes and Scholars: A Guide to the Transmission of Greek and Latin Literature*, Oxford[3] [1968].

Schubert, P. (1997), *Quatre lettres privées sur papyrus*, ZPE 117, 190–6.

Schubert, P. (2009), *Editing a Papyrus*, in *The Oxford Handbook of Papyrology*, ed. by R.S. Bagnall, Oxford, 197–215.

Sosin, J.D. (2010), *Digital Papyrology*, URL: http://www.stoa.org/archives/1263.

Stolk, J.V. (2015), *Scribal and Phraseological Variation in Legal Formulas: The Adjectival Participle of ὑπάρχω + Dative or Genitive Pronoun*, JJP 45, 255–90.

Stolk, J.V. (2017), *Dative and Genitive Case Interchange in Greek Papyri from Roman-Byzantine Egypt*, "Glotta" 93, 180–210.

Turner, E.G. (1980), *Greek Papyri: An introduction*, Oxford[2] [1968].

Van Groningen, B.A. (1932), *Projet d'unification des systèmes de signes critiques*, CE 7, 262–9.

Van Minnen, P. (1993), *The Century of Papyrology (1892–1992)*, BASP 30, 5–18.

Vierros, M. (2012), *Phraseological Variation in the Agoranomic Contracts from Pathyris*, in Leiwo – Halla-aho – Vierros 2012, 43–56.

Youtie, H.C. (1963), *The Papyrologist: Artificer of Fact*, GRBS 4, 19–32.

Youtie, H.C. (1966), *Text and Context in Transcribing Papyri*, GRBS 7, 251–8.

Youtie, H.C. (1974), *The Textual Criticism of Documentary Papyri: Prolegomena*, London[2] [BICS Suppl. 33] [1958].

Yuen-Collingridge, R. – Choat, M. (2010), *The Copyist at Work: Scribal Practice in Duplicate Documents*, in *Actes du 26e Congrès international de papyrologie (Genève, 16–21 août 2010)*, ed. by P. Schubert, Genève, 827–34.

Giuseppe G. A. Celano
An Automatic Morphological Annotation and Lemmatization for the IDP Papyri

1 Introduction

There currently exist two main digital repositories for papyri: the well-known one of the *Integrating Digital Papyrology* (IDP) Project (http://papyri.info) and the most recent one of the project *Digital Corpus of Literary Papyrology* (DCLP) (http://www.lit-pap.info). Both corpora, which can be downloaded from *GitHub* (https://github.com/papyri and https://github.com/DCLP, respectively), contain papyrological texts encoded following the de facto standard EpiDoc schema, a subset of the TEI schema specifically designed for encoding ancient documents[1].

The notorious complexity of papyrological texts is also reflected in their digital encoding, which challenges the digital humanist in many respects. The major problem is represented by the fragmentary nature of most papyri[2]. Papyrologists work to integrate texts to the best of their knowledge: sometimes just a few characters are missing, while other times full words need to be supplied, for a text to become meaningful (and even integrating single words is often not enough to understand a text). The degree of certainty for a given integration, then, varies depending on a plethora of factors, including especially the papyrologist's expertise.

The challenge posed by text integration deeply affects the TEI/EpiDoc XML encoding of texts, where specific elements, such as the `supplied` and `unclear` ones, are used to mark up additions. In both corpora, the markup is added inline (instead of standoff)[3]. As a consequence, since fragmentation in papyri often occurs at the word level, the XML markup can break up words: this phenomenon, as is known, raises computational issues when it comes to tokenizing while attempting to preserve a link to the information contained in the markup of the original document.

In this paper, I document a first attempt to add morphological annotation and lemmatization automatically to all *Papyri.info* documents. I restrict my present contribution to this repository because this is the only one containing some texts which

---

1 Cf. Reggiani 2017, 222 ff. On the DCLP see also the chapter by R. Ast and H. Essler in the present volume.

2 See N. Reggiani in this volume, § 4.4.

3 See, for example, Bański 2010. For recent XPointer-related work see https://github.com/hcayless/tei-xpointer.js (H. Cayless) and https://raffazizzi.github.io/ coreBuilder/ (R. Vigilanti)

have also been manually annotated in the *Sematia* treebank[4]. The *Sematia* texts provide us with a gold standard which can be used to measure how well the MATE tagger[5] (trained on literary texts) and a rule-based lemmatizer are expected to fare with respect to my automatically annotated corpus, which will henceforth be referred to as the MALP corpus (= M(orphologically) A(nnotated) (and) L(emmatized) P(apyri) corpus), available at https://github.com/gcelano/MALP. The MALP texts preserve the URNs of the original documents and the line break reference system for each token. Because of the complexity of the TEI inline markup in the *Papyri.info* files, I have not attempted to create stand-off annotations[6].

It goes without saying that having the possibility to search the corpora using morphological features and lemmas will positively impact our knowledge of the texts. This holds particularly true for a corpus containing a morphologically rich language such as Ancient Greek, which cannot be easily queried using simple graphic words. Moreover, morphological annotation and lemmatization dramatically speed up the treebanking process, the annotators taking advantage of some annotation being already present and correct to a large extent.

Since the *Papyri.info* corpus currently amounts to 62,901 documents, the only doable way to add morphological annotation and lemmatization to all tokens is to rely on an automatic annotator. Currently the best performing POS tagger available for Ancient Greek is the MATE tagger, which has been trained on the literary texts of the *Ancient Greek Dependency Treebank* (AGDT)[7]. The accuracy measured is 88%[8]. Comparable accuracies have been reported for the recent UDPipe tagger[9], which however adopts an annotation schema different from that of the AGDT and the *Sematia* treebank. Lemmatization will be performed using a rule-based approach relying on the *Morpheus* morphological analyzer/dictionary[10], from which lemmas can be extracted by linking it to the papyrological texts via word forms + fine-grained POS tags.

In Section 2, I will present the *Sematia* treebank, which contains morphosyntactic annotation for some of the papyri contained in the *Papyri.info* corpus: they will serve as a gold standard to evaluate the accuracies of POS tagging and lemmatization. In

---

4 Cf. Vierros – Henriksson 2017, and M. Vierros in this volume. I am aware of the existence of morphosyntactic annotations for the Herculaneum papyri (*Philodemus Project* at Würzburg), but they are not open annotations, and therefore cannot be reused.

5 Bohnet – Nivre 2012. See also http://www.ims.uni-stuttgart.de/forschung/ressourcen/werkzeuge/matetools.en.html.

6 See Bański 2010 for the issues related to TEI stand-off annotation.

7 Bamman – Crane 2011. Available at http://perseusdl.github.io/treebank_data.

8 Celano – Crane – Majidi 2016.

9 http://ufal.mff.cuni.cz/udpipe/users-manual.

10 https://github.com/gcelano/LemmatizedAncientGreekXML/tree/master/Morpheus. See also Crane 1991

Section 3, I detail the sentence-split and tokenization of the papyrological texts. Section 4 shows how the morphological annotation has been added to each token, while Section 5 details the lemmatization process. Section 6 contains conclusive remarks.

2 A gold standard for linguistic annotation of papyri: the *Sematia* treebank

The *Sematia* treebank (https://sematia.hum.helsinki.fi/user)[11] provides us with some semi-automatically annotated papyrological texts, which can be used to evaluate the accuracies of the POS tagging and the lemmatization of the MALP corpus. The treebank currently contains 224 papyrological texts annotated for their morphology (semi-automatically) and syntax (manually) according to the *Guidelines for the Annotation of the Ancient Greek Dependency Treebank 2.0*[12], which were first designed for the annotation of the AGDT. The original texts come from the *Papyri.info* corpus, but no criterion for the choice of the kind of text is at the moment followed in the treebank.

The morphological annotation and lemmatization process are performed using the *Morpheus* morphological analyzer, which was also used to help the manual annotation of the AGDT texts. Since the MATE tagger has been trained on the texts of the AGDT, both the *Sematia* texts and the output of the MATE tagger are directly comparable without conversion.

The *Sematia* treebank contains two layers of annotation for each papyrological text: the "original" layer, which only consists in the linguistic material that has been preserved in a papyrus, and the "standard" layer, which is built on the "original" layer and integrates it with editorial work aimed to make the text intelligible. Our automatic annotation is evaluated only against the texts of the standard layer: they provide a much better input for the POS tagger, the text having been integrated with missing characters/words.

Moreover, while the *Sematia* treebank contains different annotation files for each hand recognized within a papyrological text, I provide only one tokenization per text in the MALP corpus, considering the contributions of different hands as part of the same text, on a pair of any other integration. These texts were not used for comparison purposes to simplify the evaluation process (more precisely, 17 standard layer *Sematia* texts have been excluded from the evaluation phase, i.e. the ones with the following URNs (contained in `<title/>`): o.claud.1.139, o.claud.1.148, o.claud.2.227, upz.1.7, p.grenf.2.15, upz.1.59, and bgu.3.994. Moreover, the *Sematia* texts with URN

---

11 See the chapter by M. Vierros in the present volume.
12 Celano 2014.

cpr.30.28, o.claud.1.131, and o.claud.1.135 are also excluded in that the *Papyri.info* counterpart of cpr.30.28 does not contain terminal punctuation marks allowing (automatic) sentence detection, while o.claud.1.131 and o.claud.1.135 contain Latin texts.

The number of *Sematia* texts which can be used for comparison with the ones automatically generated for the MALP corpus is therefore reduced to 92 (i.e., 112 standard layers texts – 17 – 3): they are henceforth referred to as the *Sematia* comparison corpus, which is compared with the MALP comparison corpus, i.e. a subset of the MALP corpus consisting in the exact same texts but sentence-splitted, tokenized, POS tagged, and lemmatized by using the algorithms developed for this study (available at https://github.com/gcelano/MALP).

The *Sematia* texts have been sentence-splitted and tokenized using the *Arethusa* annotation tool. Punctuation in *Sematia* texts has been partly edited manually, so the number of sentences is necessarily different from that one gets by simply sentence-splitting the original texts found in the *Papyri.info* corpus. More precisely, the *Sematia* comparison corpus contains 486 sentences, while the MALP comparison corpus 461 sentences.

The intersection between these two sets of sentences is represented by 204 sentences which are exactly the same, i.e. contain the exact same tokens (with the exclusion of the elliptical tokens manually added by annotators in the *Sematia* treebank, which are simply ignored). I will use these 204 sentences to evaluate the accuracy of POS tagging and lemmatization for the MALP corpus. Even if the comparison corpus is very small, it will provide us with a rough preliminary estimate of the quality of the POS tagging and lemmatization for the MALP corpus, which can be easily calculated automatically.

3 Sentence split and tokenization

Sentence split has been achieved using a rule-based algorithm identifying sentences on the basis of presence of the following terminal punctuation marks: the period, the semicolon (which, in Ancient Greek, has the value of a question mark), and the dot above the line (whose function is comparable to that of the colon and semicolon in English). Importantly, some ancient punctuation marks are encoded in the text as XML elements, such as `<g type="mid-punct"/>`. I have converted all `<g type="mid-punct"/>` into dots above the line, which serve here just as generic terminal punctuation marks.

In the *Papyri.info* corpus there is a huge variety of `g` elements, which are used to encode glyphs[13]. Their meaning is not always clear with respect to sentence-splitting, and so they are ignored in the present study. A refinement of sentence-splitting will

---

13 See Reggiani 2017, 252 and N. Reggiani in this volume, § 4.7.

be possible by modifying the XQuery module at the relevant points provided at https://github.com/gcelano/MALP.

The XQuery algorithm needs to also be improved to correctly distinguish periods used as final punctuation marks and periods used as abbreviation punctuation marks: e.g., a single letter followed by a period is taken by rule as an abbreviation (and so the period is not tokenized). Papyrological texts can however contain a considerable amount of Ancient Greek numbers (expressed via single alphabetic characters), which are mis-tokenized if they are followed by a period at the end of a sentence.

More in general, punctuation in papyri is a very complex matter, which should definitely receive much more attention than the one so far received both in the TEI/EPIDOC XML text encoding process and in the present study, where only 'standard' punctuation marks already present in the edited *Papyri.info* files have been taken into consideration, without any attempt of developing a more complex system for sentence detection. This has as a consequence that sentence-splitting in the *Sematia* treebank, which has been manually checked, is arguably more accurate (even if not uncontroversial).

Tokenization has also been performed using a rule-based approach. Ancient Greek graphic words (i.e. space-separated words) are commonly treated as separate tokens. There are only a few arguably rare exceptions to this statement: e.g., crasis can be responsible for the phonetic and graphic merging of an article and a noun, as in θἠμέρα (= τῇ ἡμέρα). The algorithm does not attempt to provide a solution for that (and the POS tagger itself has been trained on data where other-than-space-based tokenization is not always consistent).

Since the TEI/EpiDoc XML files contain inline markup, one major tokenization-related task is to distinguish the 'data text' from the 'metadata text'. For example, the TEI `note` element contains some additional information about a part of the text, as shown in the file o.amst.8.xml (within the *Papyri.info* repository), where, with respect to a few characters, the `<note>` element specifies "Writing perpendicular to main text". The text node of the note elements should clearly not be tokenized, not belonging to the main text.

A few TEI elements can contain textual content which has been added by an editor and can be considered as part of the text: e.g., the `<supplied>` element allows integration of missing characters. In jur.pap.36.xml, for example, the textual evidence ἐπικαλουμένη is directly followed by the element `<supplied reason="lost">ς</supplied>` to signal that a final *sigma* got lost but is necessary to properly understand the text. A more complex case is represented by the element `<choice/>`, which is designed to allow different readings in a text[14]: for example, `<choice><reg>τῆς</reg><orig>τὴν</orig></choice>` on line 11 in jur.pap.36.xml shows that two different variants of the feminine article are possible at that point of the text, being one the original text on the actual papyrus and the

---

14 Cf. Reggiani 2017, 237 and N. Reggiani in this volume, § 4.6.

other one the modern 'regularization' of the non-'standard' spelling. When tokenizing, one should integrate the variants into the text one by one.

All the details about the preprocessing of a papyrological text aimed to clean it up before tokenization are contained in the XQuery module. More precisely, the function `lp:clean-markup()` deletes those elements which do not contain data text (such as the `<note>` or `<bibl>` elements) and shows the logic followed as for those elements containing mutually exclusive data text alternatives, such as the `choice` element. The XQuery module also details the rules for the tokenization itself. One tokenization problem which has not been dealt with in the present study is the case of those words splitted on different lines, as in o.narm.3.xml, where the word κολακείας is splitted into κολα (linebreak 2) and κείας (linebreak 3). The XQuery algorithm treats the two divided parts as different tokens, in that they are identified by a different line break number (`@n` in `<lb/>`), which is taken to be a major identifying reference point for a token (i.e., a token cannot span over two linebreaks). Future work should try to address this problem.

4 POS tagging

POS tagging is the NLP task aimed to add morphological analyses to tokens. A model for the MATE tagger has been trained on the AGDT, so it is possible to tag Ancient Greek texts automatically. The model has been trained on literary texts, i.e. texts whose language is quite different from that of papyri.

The token accuracy for POS tagging using the MATE tagger has been compared to the baseline accuracy consisting in assigning to each token the most frequent morphological tag associated to it in the AGDT. As explained in Section 2, the comparison is performed using the 204 matching sentences of the *Sematia*/MALP comparison corpus as a gold standard. The total number of comparable tokens is 1,839. The word forms of the tokens of each sentence have been compared as for their POS tags and morphological features.

Baseline	MATE tagger
0.56	0.62

Tab. 1: token accuracy for POS tag + morphological features.

Baseline	MATE tagger
0.67	0.76

Tab. 2: token accuracy for POS tag only.

In order to calculate the baseline accuracy we used the most frequent tags found in the AGDT and, if a word form is not present, the first POS tag + morphological features found in the *Morpheus* dictionary for that given word form. There is no clear order of entries for each token in the *Morpheus* dictionary, so this latter POS tag + morphological features assignment can be considered random.

The tables above show that having a POS tagger both for POS tags only or POS tags + morphological features is useful to get better accuracies, even if the figures are admittedly not very high.

One explanation for these low accuracies is that the vocabulary found in papyrological texts is very different from the one found in the AGDT: the distinct values of the tokens of the 204 matching sentences contained in the *Sematia*/MALP comparison corpora are 781, of which only 361 are found in the AGDT data (used to train the MATE tagger). Similarly, the *Morpheus* morphological analyzer can recognize only 395 tokens (out of the 781 distinct values). This is clear evidence that papyrological texts contain very different vocabulary from literary texts, which deserve special attention and development of specific resources.

5 Lemmatization

An attempt of lemmatization is performed using a database consisting of entries from the *Morpheus* morphological analyzer/dictionary and the *Perseus-Under-Philologic* morphological dictionary (available but not downloadable at http://perseus.uchicago.edu), which is a refinement of *Morpheus*, with correction of its errors and addition of new entries.

A typical entry of the database is like the following:

```
<d n="500083" v="a-s---fa-#ἀμετακίνητον">
    <p>a-s---fa-</p>
    <f>ἀμετακίνητον</f>
    <l r="a--s---fa-">ἀμετακίνητος</l>
    <e>ἀμετακίνητος</e>
</d>
```

Each `d` element is identified by a word form + morphological analysis, which a lemma from the *Perseus-Under-Philologic* dictionary (in the `l` element) and/or the *Morpheus* dictionary (in the `e` element) are associated to. I extracted from this database all the lemmas of those entries whose word forms and morphological analyses correspond to those of the MALP tokens (belonging to the 204 matching sentences).

Lemmatization
0.47

Tab. 3: token accuracy for lemmatization.

The result of the comparison between the lemmas in the *Sematia* sentences and the ones in the MALP sentences (containing lemmas added using the above mentioned database) is as low as 0.47. This result seems to be due to the low accuracy of POS tagging because the lemmatization process heavily relies on it: if the morphological tag associated to a word form is wrong, then the correct lemma is unlikely to be retrieved from the database. Another reason can be the fact that the vocabulary covered by the dictionaries is not complete: however, if the *Morpheus* dictionary only recognizes 395 tokens of the distinct values of the *Sematia*/MALP matching sentences, the *Perseus-Under-Philologic* dictionary has a much better coverage, which amounts to 582 tokens.

6 Conclusions

In this paper I have documented an attempt to add morphological annotation and lemmatization to the *Papyri.info* corpus automatically. The results show that fine-grained POS tagging accuracy is 0.62, while lemmatization accuracy is as low as 0.47. The fine-grained POS tagging accuracy is likely to be explained because of the vocabulary difference/coverage between the literary texts of the AGDT, which the MATE tagger has been trained on, and the papyrological texts (the AGDT recognizes only 361 tokens of the 781 distinct tokens of the *Sematia*/MALP matching sentences). This has a direct consequence on the lemmatization accuracy, which has been performed using a rule-based approach depending on the POS tagging: if the association word form-POS tag is wrong, then it is very unlikely that the right lemma can be retrieved. In order to get better results both for POS tagging and lemmatization, a papyri-specific treebank, such as the *Sematia* treebank, should be expanded, so that it can provide enough data for the training of new models (for POS tagging/lemmatization/parsing). Even if the accuracies calculated for the MALP corpus are not high, still the corpus is expected to be useful for data exploration.

7 Bibliography

Bamman, D. – Crane, G. (2011), *The Ancient Greek and Latin Dependency Treebank*, in *Language Technology for Cultural Heritage*, ed. by C. Sporleder, A. van den Bosch, and K. Zervanou, Berlin – Heidelberg, 79–89.

Bański, P. (2010), *Why TEI Stand-off Annotation Doesn't Quite Work and Why You Might Want to Use it Nevertheless*, in *Proceedings of Balisage: The Markup Conference 2010 (Montréal, Canada, August 3–6, 2010)*, URL: https://www.balisage.net/Proceedings/vol5/html/Banski01/BalisageVol5-Banski01.html.

Bohnet, B. – Nivre, J. (2012), *A Transition-Based System for Joint Part-of-Speech Tagging and Labeled Non-Projective Dependency Parsing*, in *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (Jeju Island, Korea, 12–14 July 2012)*, 1455–65, URL: http://www.aclweb.org/anthology/D12-1133.

Celano, G.G.A. (2014), *Guidelines for the Annotation of the Ancient Greek Dependency Treebank* 2.0, URL: https://github.com/PerseusDL/treebank_data/edit/master/AGDT2/guidelines.

Celano, G.G.A – Crane, G. – Majidi, S. (2016), *Part of Speech Tagging for Ancient Greek*, "Open Linguistics" 2,393–9, URL: https://doi.org/10.1515/opli-2016-0020.

Crane, G. (1991), *Generating and Parsing Classical Greek*, "Literary and Linguistic Computing" 6, 243–5.

Reggiani, N. (2017), *Digital Papyrology I: Methods, Tools and Trends*, Berlin – New York.

Vierros, M. – Henriksson, E. (2017), *Preprocessing Greek Papyri for Linguistic Annotation*, in *Journal of Data Mining and Digital Humanities – Special Issue on Computer-Aided Processing of Intertextuality in Ancient Languages*, ed. by M. Büchler and L. Mellerin, URL: http://jdmdh.episciences.org/paper/view/id/1385. [Version 1 published in 2016]

Isabella Bonati
Digital Papyrological Editions and the Experience of a Lexicographical Database

The case of *Medicalia Online*

εἰ δέ τις τῶν ἰδιωτέων γνώμης ἀποτεύξεται, καὶ μὴ διαθήσει
τοὺς ἀκούοντας οὕτως, τοῦ ἐόντος ἀποτεύξεται.
Ps.-Hp. *Vet. med*. 2,17–8 (CMG I 1, 37,17–8 Heiberg)

1 Introduction

Texts are made of words, the meaning of the words creates the text, so that under-
standing the words means understanding the text. This assertion may sound obvious,
but it actually hides a deep truth: reaching a more exact definition of the words can
help to reach a more exact meaning of the whole passage in which they occur. Hence,
the need for creating tools specifically conceived to study the words and their tight
connection with the texts. In the digital and internet-led era we are living in, elec-
tronic technologies have been profitably applied to the Humanities and intersect with
them in the scholarly field of the Digital Humanities (DH) to such an extent that it is
hard to disagree with Jerome McGann's incisive words:

> As with the Renaissance sped forward by the printing revolution of the fifteenth century, digital
> technology is driving a radical shift in humanities scholarship and education. The depth and
> character of the change can be measured by one simple but profound fact: the entirety of our
> cultural inheritance will have be reorganized and re-edited within a digital horizon.[1]

1 MᴄGᴀɴɴ 2010, 1. For a good overview of the issues entailed by Digital Humanities, cf. Sᴄʜʀᴇɪʙᴍᴀɴ
– Sɪᴇᴍᴇɴs – Uɴsᴡᴏʀᴛʜ 2004.

In this context, online dictionaries and reference tools, as results of an e-lexico-graphic process,[2] may acquire a special relevance and potential, notably if these re-sources are linked to a corpus of digital editions of texts. This is precisely the case with *Medicalia Online*, a digital lexicographical database of technical terms attested in the Greek medical papyri.[3] *Medicalia Online* is indeed strictly connected to the dig-ital editions hosted in the *Corpus of the Greek Medical Papyri Online* (CPGM),[4] the dig-ital library of ancient medical texts on papyrus recently merged into the *Digital Corpus of Literary Papyrology* (DCLP).[5] Like the CPGM project, *Medicalia Online* has been de-veloped at the University of Parma (Italy), from 2014 to the end of 2016, in the frame-work of the ERC project DIGMEDTEXT funded by the European Research Council (Grant Agreement no. 339828) and directed by Professor Isabella Andorlini.[6] Thus, given the close interconnection with the core database, *Medicalia Online* can be con-sidered as both a supplement to and an expansion of the digitized corpus of the Greek medical papyri, as I will illustrate below.

1.1 A matter of definition

Medical papyri represent a corpus of peculiar texts with a peculiar nature, that ranges from literary texts, notably treatises by known authors and *adespota*, since papyrus fragments not rarely preserve works more or less of the same status as the medical literature transmitted in medieval manuscripts, to technical texts conceived to con-vey technical knowledge, for instance technical handbooks, collections of recipes,

---

2 To explore the state of the art in the field of (e-)lexicography, see TARP 2008, FUERTES-OLIVERA – BERGENHOLTZ 2011; FUERTES-OLIVERA 2013; FUERTES-OLIVERA – TARP 2014.
3 Cf. http://www.papirologia.unipr.it/CPGM/medicalia/vocab/index.php. See also the following contributions: BONATI 2018a and 2018b, as well as REGGIANI 2017, 129–30 and 275.
4 On the CPGM Online, see the contribution by N. Reggiani in this same volume. Cf. also REGGIANI 2015 and 2018a. For a full insight of a typical CPGM digital edition, see BERTONAZZI 2018. For other references on digital papyrology and online resources and project, see in particular: ANDORLINI 1997a, 1997b, as well as RAMSAY 2004; DEL CORSO 2007; MAGNANI 2008; ANDORLINI – REGGIANI 2012; DELATTRE – HEILPORN 2014; DEPAUW – GHELDOF 2014; REGGIANI 2012, 2017 2018b; SVENSSON – GOLDBERG 2015.
5 Cf. http://www.litpap.info. Cf. REGGIANI 2017, 251–3.
6 Main reference website: http://www.papirologia.unipr.it/ERC. Cf. REGGIANI 2017, 256 and 273–5. The very first steps in the creation of the *Medicalia Online* database date back earlier, when I spent a research stay at the University of Oslo (from August 2012 to April 2013), supported by an Yggdrasil Grant from the Norwegian Research Council. There, I had the pleasure to work in close collaboration with Prof. Anastasia Maravela, who gave fundamental suggestions and a great contribution to the development of the current layout of the entries. During that first phase, I focused on the vocabulary of some representative Greek medical containers (on the topic see BONATI 2016a), which I used as samples to test advantages and disadvantages, usefulness and usability of the lemmas in *Medicalia Online*. Beside me and Prof. Maravela, other contributors in the project have been Prof. Isabella An-dorlini, Dr. Nicola Reggiani and Dr. Francesca Bertonazzi.

school manuals and catechisms, to proper documentary texts, such as public physicians' reports, petitions of private individuals and private correspondence concerning matters of health and diseases. So, besides the strictly literary texts, such a corpus mostly includes texts with a 'borderline' character, viz. combining features of papyrus documents with issues proper of the technical nature of medical writings, that are categorized – as in a sort of 'twilight zone' – as 'paraliterary' or – in a vaguely pejorative way – 'subliterary'.[7] Due to this complex and stimulating textual situation, it became clear since the beginning that a dictionary of short definitions of terms would not have fitted the exegetical requirements of ancient medical discourse. This entailed the necessity to broaden the goal of *Medicalia Online* to produce a rigorous and detailed reference collection of relevant lemmas critically discussed. An extensive and diachronic treatment of ancient Greek medical terms was indeed still missing from the scholarly landscape. It was decided to focus the attention on a selection of *specimina* with the aim of providing not merely brief explanations of many words, like in an ordinary 'dictionary', but a series of in-depth studies on selected terms. As a consequence, it is preferable to define *Medicalia Online* not as a simple 'glossary' or 'dictionary', but rather as a 'lexicographical tool'[8] containing entries or articles with an encyclopaedic flavour, or, even more specifically, as a 'specialized lexicographical tool', being it devoted to the specific set of linguistic and factual elements of the specialist subject field of ancient medicine.

Considering the vastness of the lexical material at disposal, since the papyri are a treasure-trove of linguistic information, the lexicographical process is still ongoing and potentially never-ending. Due to this aspect, *Medicalia Online* may fall into the category of the lexicographical tools "under construction" or better "dynamic",[9] according to the terms "Ausbauwörterbuch" ("dictionary under construction") vs. "Abschlusswörterbuch" ("completed dictionary") introduced by Schröder 1997, 60, and "dynamisches Wörterbuch" ("dynamic dictionary") vs. "statisches Wörterbuch" ("static dictionary") preferred by Lemberg 2001, 81. This means that *Medicalia Online* is not "a fixed object", but a flexible entity, "an organic changing database"[10] that can

---

7 Cf. Reggiani 2017, 78.

8 For a definition of 'lexicographical tool', used instead of 'reference work' to express "a superior concept for both printed and electronic dictionaries", see Tarp 2008, 123: "a lexicographical tool is a tool that can be used via consultation or passive searching by users with a specific type of communicative or cognitive need to gain access to lexicographical data, from which they can extract the type of information required to cover their specific needs".

9 On computer-lexicographical process for online dictionaries under construction, cf. Klosa 2013, 519–22.

10 Prinsloo 2001, 141.

be continually enlarged, as well as that its lexicographical process is an "open system".[11] Such a flexibility and the possibility of a constant improvement and updating represent the undeniable advantage of an online publication. Nevertheless, most of the entries already published online in the *Medicalia Online* database are going to be published also in print, in form of a collection of lexical studies, as part of a volume.[12]

1.2 Methodology and aims

There are some keywords characterizing the methodology of *Medicalia Online*. The first and most important one is 'interdisciplinarity'. The significance of interdisciplinarity in the new trends of Papyrology has been stressed several times in recent years. Suffice it to remember how often expressions like "broader concept" and "broader view", as well as words like "combination" of sources, and, of course, "interdisciplinarity" occur in *The Oxford Handbook of Papyrology* edited by Roger Bagnall in 2009. So, whilst the core focus of *Medicalia Online* is papyrological and the evidence of medical papyri plays a leading role, a systematically interdisciplinary approach inspires the inner nature of its lexical studies. This contributes to broaden the horizon of the database to a wide range of perspectives, since it provides at the same time a papyrological, linguistic, archaeological and historical-scientific overview of the studied items. Such a methodology, indeed, involves a strong sense of dialogue and cooperation among disciplines, merging and combining components of several subject areas beside papyrology: classics and history of textual transmission, digital humanities, linguistics, epigraphy, archaeology and material culture, history of science and of medical practices across the ages. It entails a critical analysis and a comparative examination of all the typologies of sources on which the study of the ancient medicine draws upon, from the written ones, i.e. papyri, literary passages (first and foremost works on medical topic, but also any other author in which the terms appear), inscriptions and *tituli picti*, to the available archaeological discoveries attesting to medical practice. Ultimately, this integrated approach, which bridges together the main subject areas in ancient studies, enables us to throw new light on the complex and multicultural setting of the Greco-Roman medicine in Egypt, and presents – to

---

11 Cf. Klosa 2013, 519: "producing an online dictionary may begin before the phase of writing is finished: online dictionaries can be published step-by-step. Thus, all phases of the computer-lexicographical process (planning – writing – producing) merge, giving yet unknown flexibility to the lexicographer. [...] While other lexicographic processes lead to an end (i.e. the publication of the dictionary), theoretically, working on an online dictionary under construction could go on forever. An online dictionary under construction is an open system".

12 Cf. Bonati – Maravela 2018.

borrow Vivian Nutton's words – "an inclusive model for understanding the medical word of Antiquity".[13]

The second keyword of *Medicalia Online* is 'verticality'. Such a comparative and thorough, i.e. 'vertical', approach to the sources contributes to improve our understanding of the Ancient World with its textual and concrete aspects, with its *verba* and its *realia*, and promotes an essentially 'vertical' rather than 'horizontal' dimension in investigating lexical items. Thus, this sort of 'archaeology of the words' also translates into the effort to 'revitalize' the past, making it more accessible to the present time.

Furthermore, starting from the evidence of the papyri, particular attention is devoted to the evolution and survival of the examined words into the modern languages and contemporary scientific discourse. Thus, one of the goals of *Medicalia Online* is to focus on the diachronic, often problematic developments of the Greek technical vocabulary tracing its trajectory from antiquity to modern times.

A further aspect concerns the analysis of the relationship, viz. the points of divergence and contact, between the terminology attested by the papyri of medical content and the – often more sophisticated – technical language known through the medieval manuscript transmission of the ancient medical writers, from the Hippocratic authors to the compendiasts of Late Antiquity. In this view, the lexical studies of *Medicalia Online* allow us to explore the contribution of the papyri to our knowledge of the Greek medical language.

1.3 The database and the entries

The database is built on the open source vocabulary server *TemaTres* and is browsable in different ways. The home page displays a threefold subdivision by macro-categories, each of which is further divided into subcategories providing a taxonomical classification of the terms: "Lexicalia", i.e. word typologies (e.g. containers, ingredients, instruments, *termini technici*), "Medical branches" (e.g. gynecology, ophthalmology, pathology, pharmacology, surgery), "Text typologies" (e.g. *adespota*, catechism, documentary texts, prescription). A single term can also be subordinated to two or more subcategories, so that it can be searchable in each of them. Another way is to browse the terms and the categories alphabetically by clicking a certain Latin or Greek letter either at the top or at the left bottom of the home page. Finally, on the very top, it is also possible to use a full-text search, as well as an "Advanced search". In the latter case, a drop-down menu provides a submenu of navigation items to select the research scope: "Term" restricts the search to the headwords, "Meta-term" to the

---

13 NUTTON 2004, 16.

categories, "Non-preferred term" to secondary headwords such as variants and diminutives, and "Note" to the thematic boxes.
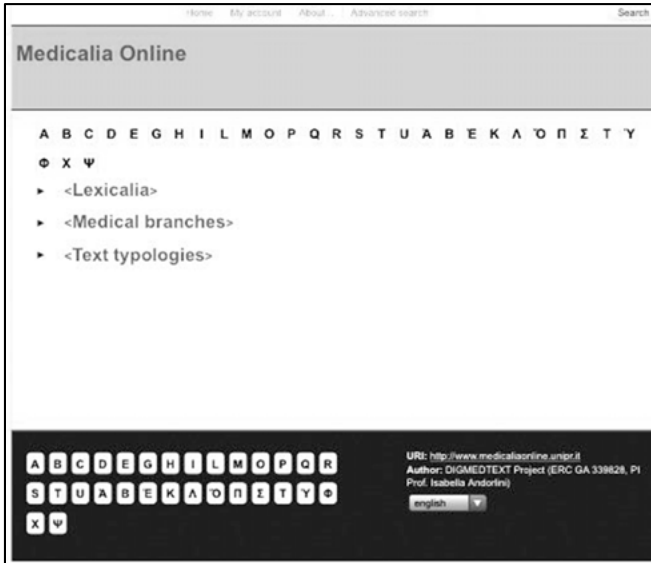
Fig. 1: The database home page.
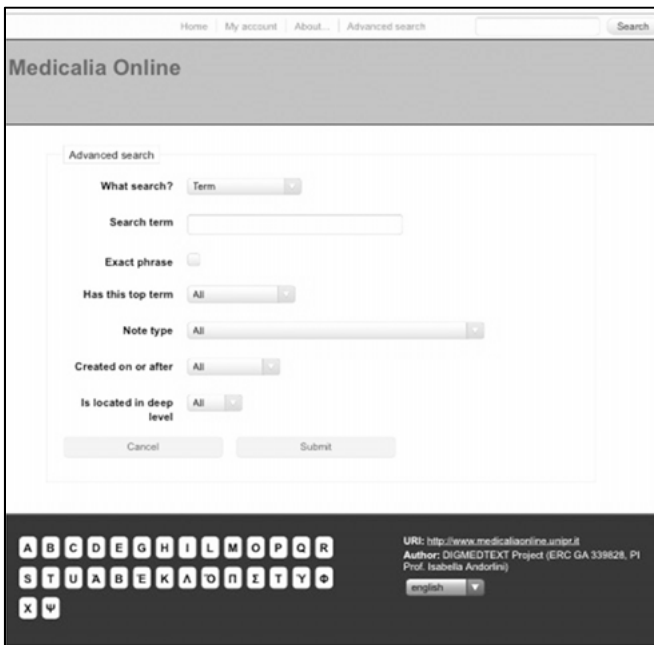
Fig. 2: The advanced search inferface.

The lexicographical structure of the lemmas is innovative and reflects the interdisciplinary, integrated approach that inspires *Medicalia Online*. The layout of the entries is conceived to offer a broad overview of the examined words and is essentially comprehensive, but, at the same time, it is user-friendly and applicable to any lexical category and semantic field. User-friendliness is indeed an important prerequisite when a lexicographical tool is complex and involves a conspicuous bulk of information. Each lexical entry consists of a fixed set of thematic boxes ("notes"), as follows:

– "Variants" includes a list of variants, both grammatical (e.g. diminutives) and phonetic/spelling variants as found in the papyri, the Latin transliteration or form(s) of the term, and the cognates of medical relevance, if any;

– "General definition" gives a dictionary-like definition useful to provide the reader with the main information concerning the searched term and its 'immediate' meaning before (or in case) (s)he goes on reading through the full lemma;

– "Language between text and context" is a linguistic section containing discussions on etymology, morphology, semantics, variants and cognates of the examined term, but it also discusses its linguistic history up to modern times and the diachronic developments of its technical meaning(s);

– "*Testimonia* – a selection of representative sources" lists some Greek and Latin passages from all kind of written sources (literature, papyri, inscriptions) in which the term is attested, selected according to their medical relevance. Each passage is accompanied by an English translation;

– "Commentary" is the most substantial section of the entry and is aimed at contextualizing the term in its textual and historical-scientific background. In order to do this, the section is divided into two chapters. The first one ("[the term] and its medical sources") traces a detailed overview of what the ancient sources attest about the term, also scrutinizing the possible changes of its semantic value over time from its earliest attestations to Late Antiquity, and the comparison between its ancient and modern meaning(s). The second chapter ("[the term] in practice") is specifically focused on the 'practical' side of the examined item and outlines the connection between the word and its concrete dimension. To make just some examples, this means the material reconstruction of the related object in case of words denoting *res medicae*, such as containers employed to prepare or store remedies or surgical implements, and the methods of treatment and surgical procedures to be performed when dealing with names of pathologies and disorders, with particular attention to parallels, divergences and innovations along the history of medicine;

– "Bibliography" includes "Lexicon entries", i.e. dictionaries, glossaries etc., and "Secondary literature", i.e. more extensive studies on that particular topic or word.

– "CPGM/DDbDP reference(s)" lists the papyrological evidence containing the word. Since some of the examined terms occur only in the CPGM Online, while others appear also or only in documentary texts dealing with medical topics, such

as private letters requesting remedies or surgical instruments, which are contained in the DDbDP, "the documentary evidence ('DDbDP references') will be linked to the appropriate texts on *Papyri.info*, the literary or paraliterary one ('CPGM references') to the forthcoming texts on DCLP, from which, in turn, it will be possible to insert links back to *Medicalia Online*".[14]

Finally, a clickable list of the terms connected to the main term (diminutives, variants, cognates, Latin forms), which are also searchable through the alphabetical list that can be found both at the top and at the bottom of the home page, closes the lemma.

2 The lexicographical database and the digital editions of texts

2.1 The interconnection between *Medicalia Online* and the textual database

The contribution of Greek and Latin papyri to our knowledge of classical languages is an indisputable fact that has been scholarly recognized since their discovery in the dry sands of Egypt in the late 19th century. It is worth quoting EVANS – OBBINK 2010, v:

> Every scrap of papyrus and every ostracon or tablet unearthed has the potential to change some aspects of the way we think about these languages. Such texts have the capacity to modify our understanding of the classical forms of both languages and for their post-classical development provide evidence of the most direct kind we shall ever acquire. The richness of the resource can hardly be overstated.

Exactly like the other categories of papyri, papyri of medical content have a massive linguistic potential. The corpus of the Greek medical papyri has indeed not only enhanced our knowledge of medical literature and everyday medical practice, revealing valuable information on the diseases that affected people in the Egyptian *chora*, as well as their pharmacological and surgical treatment. It has also offered rich attestation of Greek technical vocabulary, its diachronic trajectory over time, its registers and levels of technicality, from the actual medical Greek written or spoken by medical professionals when communicating with their colleagues, to the not properly technical but still medical language used in everyday life by lay persons and practising physicians.[15]

---

14 REGGIANI 2017, 130.

15 On the contribution of medical papyri to the study of medical Greek, cf. MARAVELA 2017.

In addition to the new attestation of technical terms already well known by means of the 'official' medical writings transmitted through the medieval manuscript transmission, medical texts on papyrus often refine our knowledge of weakly attested or extremely rare words and bring back to light elements of Greek medical vocabulary previously lost and completely unattested in other medical sources.

An in-depth study of medical micro-language and technical terminology is thus essential to deeply understand the texts. For this fundamental reason, to join the digital editions of the medical texts on papyrus with the lemmas in *Medicalia Online* can be of the utmost importance to promote an integrated and mutual enrichment.

An even more in-depth investigation might be realized by adding to the lexical studies the analysis of the morpho-syntactic stucture of the texts by means of the application of different levels of linguistic annotation. Annotation is indeed a cardinal part of the linguistic analysis of a corpus of texts, the method of describing, recording and analysing linguistic phenomena through computer-based text corpora that is better known as Corpus Linguistics.[16] As stressed by REGGIANI 2016, 2:

> 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.

Thus treebanking, as it is used in linguistics, is a possibility to model how sentences are built by creating morpho-syntactic trees. In the field of Classics this kind of linguistic annotation is now at a very advanced level.[17] Just to mention two relevant projects, *The Ancient Greek and Latin Dependency Treebank* (AGDT 2.0) is a corpus of ancient Greek and Latin literary works, annotated on the morpho-syntactic and semantic layers, which has been developed since 2006 at the Leipzig and Tufts Universities by Giuseppe G.A. Celano, Greg Crane, Bridget Almas and others,[18] while, on the more strictly papyrological side, the project *Sematia*, conducted by Marja Vierros and Erik Henriksson at the University of Helsinki, is a platform aimed at facilitating

---

16 On this issue, see for example BIBER – CONRAD – REPPEN 1998, FACCHINETTI 2007 and KUEBLER – ZINMEISTER 2014, as well as the chapters by N. Reggiani and M. Vierros in this volume.

17 Cf. REGGIANI 2017, 180–6 with bibliography in n. 13.

18 See at https://perseusdl.github.io/treebank_data. See also G. Celano's chapter in this volume.

the linguistic tagging of digitized documentary papyri through the creation of linguistic layers from TEI/EpiDoc XML documents.[19]

It is worth mentioning another innovative text analysis tool, which might be potentially useful also in the case of the corpus of Greek medical papyri. I am referring to CATMA 5 (*Computer Aided Text Markup and Analysis*),[20] a tool developed at the University of Hamburg that offers an interesting combination of three main features, since it allows collaborative annotation and analysis of a text or a text corpus, it supports explorative, 'non-deterministic' practices of text annotation, viz. a discursive and debate-oriented approach to text annotation based on the research practices of hermeneutic disciplines, and integrates text annotation and text analysis in a single web-based working environment.
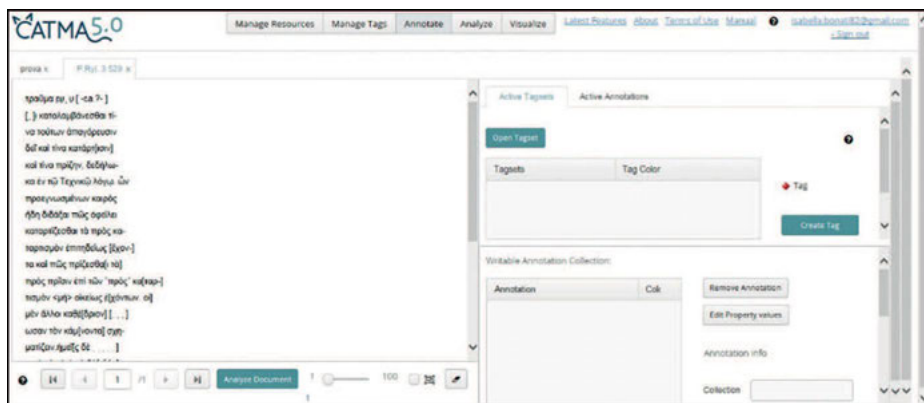
Fig. 3: CATMA screenshot

One of the main outcomes provided by the digital tools is the possibility to support several kinds of linguistic analysis – lexical, semantic, morphological, syntactic – in direct interconnection with the texts, namely directly *on* the digitized textual editions. This allows an – even simultaneous – work of in-depth investigation, abstraction and conceptualization *on* and *through* the text itself, thus enhancing its deep comprehension and interpretation in an immediate, dynamic and interactive way. Immediacy, dynamism and interactivity are indeed among the most stimulating features and perspectives of a digital publication. In the case of *Medicalia Online*, its interconnection

---

19 See at https://sematia.hum.helsinki.fi. See also M. Vierros' chapter in this volume
20 See at http://catma.de. Thanks to the funding made available by the North-West University of Potchefstroom, I had the opportunity to be introduced to CATMA during the workshop "Digital Annotation and Analysis of Literary Texts. A hands-on introduction to CATMA", organized by the South African Centre for Digital Language Resources (SADiLaR) and held by Prof. Christoph Meister at the University of Pretoria (August 21, 2017).

with the core textual database may exactly be defined with these adjectives: immediate, dynamic and interactive. The expected systematic connection of the examined words occurring in the digital editions of the medical papyri to the related lexical entries in *Medicalia Online*[21] may help not only to refine the definition of a term in a particular papyrus text, but also to contextualize the term in all the other sources attesting it. *Medicalia Online* is, thus, so conceived as to integrate the textual database, in order to serve as an expansion and a supplement to the digital editions of the Greek medical papyri.

The issue of the integration and interconnection with the main database puts the stress once again on the contribution of a close lexicographical examination to an enhanced textual understanding. In particular, a thorough lexicographical approach can strengthen a double awareness of the examined lexical item within the framework of the text, as it will be illustrated by means of some selected case-studies: 'concrete' on the one hand, 'textual-philological' on the other hand.

2.2 Lexical studies and 'concrete' awareness

As already said, the interdisciplinary and critically comparative methodology of *Medicalia Online* enables us to explore also the practical and material dimension concerning the term and its context. This aspect makes it possible to almost physically 'visualize' the concrete reality under and beyond the words, especially thanks to the second section of the "Commentary" (C 2), which is explicitly focused on the concrete side of the lexical items (see § 1.3). Thus, this sort of 'archaeology of the words', aimed at 'revitalizing' the past, makes the ancient texts themselves more 'living' and their words almost 'tangible'.

An example of an object virtually 'reconstructed' starting from the evidence of a papyrus is provided by the word ὑδρία, commonly denoting a particular type of container used as a jar to carry and pour water.[22] This primary function, made clear by the etymology from ὕδωρ, "water", is well established in the ancient sources. In the documentary papyri, this container is also filled with other contents, such as foodstuffs.[23] Only two passages of medical authors refer to the ὑδρία (and to the diminutive

---

21 For the time being it is only possible to create links between the lemma (see section E, "DDbDP/CPGM references" mentioned above, § 1.3) and the words highlighted in the front matter of the digital editions in DCLP, as well as, conversely, between the latter and the corresponding lemmas in *Medicalia Online*, but it is expected that in the near future it will be also possible to insert links directly into the Greek texts both in DCLP and in *Papyri.info*, in case of medical terms found in documentary papyri of medical content.

22 Cf. *MedOn s.v.* and Bonati 2016a, 157–75.

23 Cf., e.g., PSI IV 428,89–90 and 92 (pomegranates), BGU XIII 2359,10 (beans), P.Oxy. I 155,4 (bread).

ὑδρίσκη) as a container of pharmaceutical use.[24] In both cases, the vessels have the function of small containers for ointments employed for the storage of the therapeutic products prior to their use. This function is attested in just one documentary papyrus of medical content dating back to the late IV century CE, P.Oxy. LIX 4001. The papyrus is a letter written by a certain Eudaemon to his mother, grandmothers and a woman called Cyra. Eudaemon, who is a doctor, sends the letter to his surgery presumably in Oxyrhynchus, as the address on the back shows (ἀπόδος εἰς τὸ ἰατρεῖον). He is working away from home for professional reasons, and he now asks for the means to make some medical implements on his own. He also notifies his family that he received a "*hydria* of eye-salves" instead of a "*hydria* of animal grease".[25] The passages of medical content do not provide any information about the physical appearance of the vessel, but the word ὑδρία corresponds to a well-recognized type of container in the (conventional) archaeological vocabulary, and it is one of the most common shapes depicted in Attic vase-painting, such as in the famous François vase (ca. 570 BC). According to all types of ancient evidence, the most typical features of this vessel are the presence of some handles (usually three: two horizontal side-handles for easy lifting and a vertical handle for pouring the water or carrying when empty) and a fairly narrow neck set off from the body. It is significant to stress that the comparative study of all the sources – written as well as archaeological – allows us to conjecture about the material shape of the ὑδρία in the Oxyrhynchus papyrus, as well as, more broadly, in the other two medical passages, even though it does not appear any explicit information about its morphology. First of all, although in these texts of medical content ὑδρία / ὑδρίσκη does not represent an actual *terminus technicus* in the vocabulary for medical containers, it is likely that the word has been used in connection with remedies having an ointment-like consistency because of the shape and the considerable versatility and manageability of this small vessel. Assuming that the ὑδρία / ὑδρίσκη of the medical sources has a narrow neck and a vertical handle like some glass *hydriskai* for cosmetics and oils found in archaeological contexts,[26] its morphology seems

---

24 Cf. Gal. *Comp.med.loc.* I 2 (XII 437,2–5 K.) and Paul.Aeg. III 2,2,4–6 (CMG IX 1,132,19–21 Heiberg).

25 Cf. ll. 22–30 ἔσχαμε(ν) | δὲ καὶ τὰ ἄλλα πάντα χωρὶς μόνης | τῆς ὑδρείας (*l.* ὑδρίας) τοῦ ὀξυγγείου (*l.* ὀξυγγίου). ὅθεν | σπουδασάτω ὁ ἀδελφὸς ἡμῶν | Θεόδωρος ζητῆσαι ηπο . . . . τον | να . . . . . καὶ γνῶναι περ[ὶ] αὐτοῦ | [. .] . . . . . . ὑδρείαν (*l.* ὑδρίαν), παρέσχεν ἀν|τὶ τοῦ ὀξυγγείου (*l.* ὀξυγγίου) κολλουρίων ὑδρε¡[⟦ . . ⟧αν (*l.* ὑδρίαν): "we had all the other things too except only the jar of grease. So let our brother Theodorus be eager to search for it … [make sure to look …] and to know about it … jar, he provided instead of the grease a jar of ointment".

26 For these core-formed glass miniatures (*hydriskai*), imitating the shape of the three-handled and narrow-necked pottery *hydria* and very probably containing scented oils or cosmetics, see Bonati 2016a, 172 and Stern 1999, 29–39.

to be particularly suitable for closing and sealing.[27]C:\Users\Isabella\Dropbox\Medicalia Parma\HYDRIA-final version.docx Thus, it is highly probable that these features represent the main reason why in the papyrus letter from Oxyrhynchus a little vessel named ὑδρία has been chosen to contain the therapeutic products mentioned: the κολλούρια actually received by Eudaemon and the ὀξύγγιον previously requested by him but never dispatched. In all likelihood, these special details of the ὑδρία attended to by Eudaemon's family will have assured proper preservation of the remedies during their transport from the ἰατρεῖον in Oxyrhynchus to the village where Eudaemon was apparently working as a physician.

A further example of this attempt to 'revitalize' and 'reconstruct' an object from the past may be represented by another name of container, φαρμακοθήκη.[28] The term is a semantically transparent compound employed to denote a portable chest or case for the storage of remedies and medical implements. The earliest witness of the noun is P.Oslo II 54,6 (second half of the II – first half of the III century AD), a private letter on papyrus from Egypt (likely coming from the area of Oxyrhynchus), addressed by a certain Horeion to his father Apollonios. The term occurs otherwise exclusively in astronomical and Christian works from the V century AD onwards.[29] The lack of this term from the medical texts is significant and might suggest that, even if the compound φαρμακοθήκη indicates an exclusively medical container, viz. a technical accessory, it never developed into an actual *terminus technicus*. It is likely that it served as a lay synonym for the physician's tool-case in common language, for instance in practitioners' everyday conversations with their patients, instead of more 'official' and technical terms for similar objects used among professionals, such as νάρθηξ,[30] a noun which is well known in literary sources, and explicitly defined νάρθηξ ἰατρικός in a literary papyrus containing alchemical recipes (PSI inv. 22011,48 = TM 65816), but

---

27 This aspect finds a confirmation in the fact that some ὑδρίαι ἐσφραγισμέναι used for transporting and sending products are mentioned in two documentary papyri, SB X 10559,1 (V AD, ?) and CPR XXV 25,4 (VI–VII AD, Arsinoites or Herakleopolites).

28 Cf. *MedOn s.v.*, BONATI 2016b, 663–7 and 2016a, 185–95.

29 Cf. e.g. Cat.Cod.Astr. I 104,26–30 Olivieri and Procl. *Or.* XVIII *in laudem apostoli Pauli* (PG LXV 817D–820A Migne).

30 It refers to a narrow cylindrical type of case, named after the *Ferula communis* (νάρθηξ in Greek), with the wood of which it was originally made. This name continued to indicate such cases even when made of other materials, and came to designate boxes with different shapes and functions. For a discussion on the term and for references to ancient sources and to the bibliography, see especially MARGANNE 2004a, 122–4.

also – though less attested – δελτάριον,[31] ἐγχειρίδιον,[32] and πήρα.[33] As a matter of fact, according to SCHIRONI 2010, 338:

> Technical terms often have lay synonyms in common language; this is particular evident in medicine where technical and lay terminology coexist [...] and often physicians use the latter in order to be understood by the patients.

Thus, it seems that the compound φαρμακοθήκη might be placed at an intersection point between the language of everyday life often documented in the papyri and the technical vocabulary for medical containers. Furthermore, even though the extant written sources do not provide illuminating information about the shape and the physical appearance of φαρμακοθήκη, archaeological evidence may represent a fertile ground to start formulating hypotheses. Indeed, a certain number of cases for storage and transport of drugs and remedies and/or surgical implements have been unearthed in excavations. Among these, a characteristic typology is rectangular (on average 12 × 6–7 × 2–3 cm), equipped with a sliding outer lid and internally divided into compartments, each having its own hinged cover in order to store together different medical substances with no risk of contamination. Assuming a terminological and typological overlapping between the φαρμακοθῆκαι mentioned in the written sources and any of the archaeological *boîtes médicales*, it is tempting to surmise that this compartmentalized rectangular type with sliding lid was the most suitable shape for the domestic pharmacy chest mentioned in the Oslo papyrus. This hypothesis relies on the association between the almost intuitive formation of the compound – literally a case for medicines (φάρμακον + θήκη) – and the most common type of container with that function in both professional and non-professional contexts. Archaeological discoveries have offered several *specimina* dating back to the Roman period, therefore contemporary with the papyrus from the Oslo collection. Often very well preserved, and sometimes still holding residues of their pharmaceutical contents, some of these θῆκαι surely derive from a professional field, but a similarity in shape of their household counterpart may be likely presumed.[34]

---

31 δελτάριον occurs with this meaning only in P.Oxy. LIX 4001,30–1. For a discussion on the term and for references to ancient sources, cf. ANDORLINI 1996, 7–8 (especially n. 5); FISCHER 1997; MARGANNE 2004a, 124–5 and 2004b, 23–4 with bibliography, as well as pp. 31–3 (with n. 30) and pp. 38–40.

32 ἐγχειρίδιον is a case – as the etymology suggests – suitable to be carried "in the hand", likely similar in form to a δελτάριον, cf. MARGANNE 2004a, 125–6.

33 πήρα is listed in GMP II 10,6–7 (= P.Strasb. Copte inv. 563,6–7) τὴν πήραν | τοῦ ἰατροῦ with other medical instruments and glossed with the word ἐγχειρίδιον by Ammon. *Diff.* 390,2 (101,5 Nickau), cf. MARGANNE 2004a, 126 and FISCHER BOVET 2009, 163–4.

34 For some significant examples of this type of rectangular boxes for medical purpose, see in particular BLIQUEZ 1994, 69 and 191 no. 296 (ills. 189–90); DENEFFE 1893, 37–8 with Pl. 2 (ills. 1 and 6); MILNE 1907, 172–3 with Pl. LIV; KÜNZL 1996, Abb. XXXIV.

Finally, in both of these illustrative cases, an operation of 'archaeology of the words', consisting in integrating written and material evidence, contributes to shed new light on the (hypothetical but plausible) physical reality hidden behind the ancient words.

2.3 Lexical studies and 'textual-philological' awareness

A thorough analysis of a term in all its written attestations can bear huge potential by improving our awareness of the ancient texts and their – sometimes highly problematic – philological issues. In-depth lexical studies like those offered by *Medicalia Online* can indeed give a considerable contribution to promote the contextual and philological interpretation of medical texts, sometimes even resulting in new readings and corrections to the established editions. The critical exegesis of these texts is strictly dependent on the understanding of their technical terminology, to an extent that reaching a more exact definition of the words expressing the contents can enhance the philological awareness of the textual context in which they occur.

Furthermore, *Medicalia Online* allows both a 'holistic' and an 'atomic' approach to the object of study. The analysis is carried out in a 'holistic' way in the sense that the term is studied in all its occurrences in order to get the broadest possible overview of its meaning(s). At the same time, it is also performed in an 'atomic' way, so that the word is grasped in its verticality and all its details are accurately examined and understood. This operation encourages a 'double movement': from the text(s) to the word – the 'holistic' study of the term in all the written sources attesting it leads to a deeper comprehension of the term itself –, and from the word to the text(s) and the textual tradition. This thorough knowledge of the word entails a better understanding chiefly of the medical text on papyrus that is the starting point of the lexical analysis, but also, very often, of the passages – or at least of some of them – by medical authors that use this word. This implies a more refined philological awareness that can also enable us to amend errors in the manuscript transmission.

To illustrate the contribution of an in-depth analysis of the technical terms to the exegesis of the medical texts – whether medical papyri or passages by medical writers – it may be useful to take as a *specimen* the observations raised by the study of the word καθέδριος.[35] καθέδριος is an adjective literally meaning "of or for sitting" (LSJ⁹ 851 *s.v.*). It is derived from the name of the object that concretely receives the action

of sitting, καθέδρα ("seat"). As to the word formation, καθέδριος is a denominative adjective in −ιος, suffix of Indo-European origin usually employed to form adjectives from noun-stems and productive during the entire history of ancient Greek, from the Homeric language to the *Koine*.[36] Like all the other derivatives of the noun καθέδρα,[37] καθέδριος seems to be quite late, since it does not appear before the II century AD. Its earliest extant attestations are two medical texts on papyrus, and this is the reason why the evidence of the papyri is particularly relevant in this case: P.Aberd. 11,10, an ophthalmological catechism of the II century AD (http://litpap.info/dclp/63332), and P.Ryl. III 529r,57 (and maybe ll. 70−1), a fragment from a papyrus codex of the late III century AD containing a surgical treatise concerned with the treatment of shoulder dislocation (http://litpap.info/dclp/59970).

Starting from these early attestations, καθέδριος is an adjective appearing only in medical sources to denote the sitting position of the patient, so that it seems to be a genuine *terminus technicus* of the medical micro-language. Its technical value is strongly confirmed by the medical authors, from Oribasius to Paul of Aegina, where the adjective is mostly used in the description of different kinds of surgical operations, but also in other medical contexts. The common formula is καθέδριος + ὁ κάμνων / πάσχων + σχηματίζω vel sim., with καθέδριος, as a rule, in predicative position,[38] but the adjective is also associated with the noun σχῆμα.[39]

In order to contextualize the topic, it might be relevant to stress that placing the patient in the proper position has been an important prerequisite for surgery since ancient times. Like nowadays, different physical positions were required for different procedures. Therefore, it was essential to identify the correct position for any given operation. The importance of surgical positioning emerges, for instance, from a passage of the Hippocratic treatise *De officina medici* (*Off.* 2 = III 275−6 L.):

τὰ δ' ἐς χειρουργίην κατ' ἰητρεῖον· ὁ ἀσθενέων· ὁ δρῶν· οἱ ὑπηρέται· τὰ ὄργανα· τὸ φῶς· ὅκου· ὅκως· ὅσα· ὅκως· ὅκου τὸ σῶμα, τὰ ἄρμενα· ὁ χρόνος· ὁ τρόπος· ὁ τόπος,

---

36 Cf. Chantraine 1933, 33−8.

37 The noun καθέδρα, a compound of ἕδρα «seat, chair» (< κατά + ἕδρα), is in itself a derivative in -ρᾱ of ἕζομαι "seat oneself, sit" (LSJ⁹ 478 *s.v.*), from the IE root *sed "sit down". All the derivatives of καθέδρα seems to be of quite late formation, such as the neuter diminutive καθεδράριον "little seat", appearing in a private letter on papyrus, P.Oxy. VI 963 (II–III AD), and the compound κλινοκαθέδριον "easy chair" (LSJ⁹ 961 *s.v.*), gloss of κλιντήρ "couch" (LSJ⁹ 961 *s.v.*, cf. e.g. *Et.M.* 520,26−7 Kallierges).

38 Cf. e.g. Orib. *Coll.* XLVI 11, 2,1−3,3 (CMG VI 2,1, 219,30−5 Raeder); Paul.Aeg. VI 8, 1,10−1 and 90, 4,10−1, as well as 99, 2,1−6 and 101, 1,3−6 (CMG IX 2, 51,11−2 and 139,10−1, as well as 152,14−9 and 156,19−22 Heiberg).

39 Cf. e.g. Aët. XV 5,50−2 (19,15−7 Kostomiris) ἐν μὲν οὖν τῷ ἐνεργεῖν, σχηματιζέσθω ὁ πάσχων κατακεκλιμένος, τὸ γὰρ καθέδριον σχῆμα εἰς λιποθυμίαν τάχιστα προτρέπει τὸν πάσχοντα: "during the surgical operation the patient must be placed lying on the back, for the sitting position causes quickly the patient to faint".

in which the adverbs ὅκου and ὅκως express respectively "where and how" and refer to what precedes in the list, that is "the patient, the operator, the assistants, the instruments, the light".[40] This aspect is also evident in a medical papyrus from Oxyrhynchus, P.Oxy. LXXIV 4972, dating back to the II–III cent. AD, which belongs to the catechistic genre and contains a systematic exposition of the divisions of surgery (http://litpap.info/dclp/119317). In ll. 3–7, the most correct positions for certain operations are discussed and, in particular, it is mentioned when the patient is lying back on a sloped couch with either the head (ἀνάρροπον) or the feet (κατάρροπον) raised higher:

> τὸ δὲ] σχηματικόν ἐστιν τὸ τῶ[ν | ἐπιτη]δίων σχημάτων ὥσπερ | ὅταν] λέγωμεν ἀνάροπον (*l.* ἀνάρροπον) ἢ | κατάρ]οπον (*l.* κατάρροπον) σχηματίζιν (*l.* σχηματίζειν) τόν | κάμν]οντα

> The position-based is that concerned with appropriate positions, as when we speak of positioning the patient tilted up or tilted down.
>
> (transl. Leith 2009, 63)

The synchronous 'holistic' and 'atomic' study of the adjective καθέδριος in all its occurrences led to suspect some textual uncertainties in two passages from late antique compendiasts. The first one, transmitted by Oribasius, is part of an excerpt taken from the work entitled *Περὶ ποιουμένων βοηθημάτων* by Antyllus, a surgeon and physician of the II century AD. The passage concerns the most proper couch for the patient. It is reported that a tilted couch, i.e. a couch with the head higher than the feet, puts the patient under strain because (s)he is like seated on a chair, even if this position is suitable for those who suffer from head pain. The text printed by the Teubner editor of Oribasius, J. Raeder, runs as follows (*Coll.* IX 14,6,1–3 = CMG VI 1,2, 15,21–3 R.):

> ἡ (*sc.* κλίνη) δ᾽ ἀνάρροπος σφόδρα κόπου ποιητική, ἐοικυῖα καθεδρίου σχήματι, τοῖς δὲ περὶ κεφαλὴν οὐκ ἀνάρμοστος.

The manuscript tradition is unanimous in recording καθεδρίου σχήματι,[41] but U.C. Bussemaker and Ch. Daremberg in their previous edition of the *Collectiones Medicae* (Paris 1854, p. 310,4) make the emendation καθεδρίῳ σχήματι. In the former case, the term in genitive is a noun, i.e. the neuter καθέδριον, and is referred to the material object, a small chair.[42] But this juxtaposition is unparalleled, whereas the adjective

---

40 As to the patient, ὅκου means where and ὅκως how the patient is placed, that is his position, cf. comm. *ad l.* by Littré (Paris 1841) 276. Cf. also Gal. *In Hp. Off.* I 6 (XVIIIb 668,9–670,5 K.).

41 Cf. ThGL V 770D *s.v.* καθέδριον.

42 The only certain medical attestation of the neuter noun καθέδριον is Sor. *Gyn.* II 37,5,1–4 (CMG IV, 80,21–4 Ilberg) μικρὸν δ᾽ ἐν ταῖς ἀγκάλαις αὐτὸ προδιακατέχουσα μετὰ τὸ συμμέτρου μετασχεῖν γάλακτος κοιμιζέτω καθ᾽ οἵας ὑπεδείξαμεν κοίτης, προκύπτον δὲ καὶ ἐγκλῖνον καθεδρίῳ, where καθέδριον represents the stool on which the woman who breastfeeds the newborn is seated and bends

καθέδριος + σχῆμα is certainly attested, as aforementioned, and appears in dative also in another excerpt from the same work by Antyllus (cf. *Coll.* VI 23,9,4–10,1 = CMG VI 1,1, 181,23–4 Raeder: τῶν δ' ἀπυρέτων ⟨τοὺς ......⟩ καθεδρίῳ σχήματι αἰωρητέον). Thus, despite the consensus of the manuscripts, the study of all the attestations of the term leads to approve, on a grounded basis, the correction made by Bussemaker and Daremberg.

Also in a passage from Paul of Aegina, in a chapter on the insertion of the catheter in a male bladder, the occurrence of the term seems to conceal an element of philological uncertainty. The text edited by J.L. Heiberg (VI 59,1,9 = CMG IX 2,98,10 H.) is τὸν δὲ κάμνοντα σχηματίσαντες εἰς καθέδριον, literally "having placed the patient on a seat". According to Heiberg's apparatus, the only divergence in the manuscript tradition should be the presence of καί following καθέδριον in two codices of the XIV century from Paris (**D**, cod. Paris. Gr. 2208, and **F**, cod. Paris. Gr. 2292). However, it might be relevant that the second hand of a codex of the X century, **V** (cod. Paris. Gr. suppl. 446), deleted εἰς. This deletion might reflect the common expression with καθέδριος as an adjective (τὸν δὲ κάμνοντα σχηματίσαντες καθέδριον), whereas in the text transmitted by the manuscript tradition and printed by Heiberg καθέδριον is a noun and is unparalleled in this formula. Therefore, in case the deletion made by **V²** is correct – as it seems to be given the occurrence of the term in similar expressions –, the adjective καθέδριος would acquire its common predicative value, even if it is usually placed before ὁ κάμνων / πάσχων, and not after, as it would be in this passage.

Even more significant from the textual viewpoint are the aforementioned medical papyri attesting the word. In the earliest papyrus, P.Aberd. 11,[43] the term appears in the discussion on pterygium surgery immediately before the lacuna at l. 10 (cf. ll. 9–13):

> χειρ[ουργεία τοῦ πτερυγείου.] | μετὰ τὸν καθέδρειο[ν ὄντα τὸν πάσχοντα, ἐκ] | τοῦ ὀφθαλμοῦ διφ[υῆ βλέφαρα διαστείλαντες] | τὸ πτερύγειον δι[εκφανοῦμεν ἀγκι]|στρείοι, βελόνην [δὲ λίνον καὶ τρίχα ἱππείαν]

forwards. In such a case, the term bears no technical value, being a simple (and generic) stool. Moreover, the neuter καθέδριον occurs only twice in late non-medical texts. In Zonar. δ 524,27 Tittmann *s.v.* διέδριον the term καθέδριον is used as the *interpretamentum* of διέδριον, a compound of ἕδρα meaning a "seat for two persons" (LSJ⁹ 423 *s.v.*, cf. also *Suda* δ 896,1–3 Adler *s.v.*). It is probable that here the gloss καθέδριον simply represents a generic (though not recurring) word for a "seat", without a more specific correlation with διέδριον. In *schol.* A. *Th.* 454h (209,7 Smith) τῶν παρθενικῶν καθεδρίων· ἐδώλιον δὲ κυρίως ὁ ζυγὸς τῆς νηός, the juxtaposition παρθενικῶν καθεδρίων appears to be the explanation of *Th.* 454–5 πωλικῶν / θ' ἐδωλίων. Thus, the plural of καθέδριον – as well as of καθέδρα – seems to acquire the extended semantic value of "bower, abode", that is one of the meanings of the plural ἐδώλια (cf. LSJ⁹ 478 *s.v.*), even though the second part of the *scholium* refers to the meaning of ἐδώλιον as "rowers' benches".

43 *Ed.pr.* Turner 1939, 13. The text was republished (*ed.alt.*) in Marganne 1994, 104–11.

> Surgery of pterygium. After the patient is seated, having separated the two eyelids of the eye, we will isolate the pterygium from the eye with a hook, and a needle having a thread and a horsehair...

The restoration ὄντα τὸν πάσχοντα, with the participle dependent on μετά, "after" in temporal sense, is a likely supplement, but also κάμνοντα may be considered.[44] On the one hand, the verb εἰμί is unparalleled in this expression, on the other hand the space is not enough for the usual σχηματίζοντα, and this makes the presence of ὄντα plausible. Moreover, the article before the adjective seems to suggest an attributive value, whereas in the medical authors, as already seen, καθέδριος is always in predicative position. Different is the restoration proposed by TURNER 1939, 13: μετὰ τὸν καθέδρειο[ν βίον ?]. It was very probably influenced by a passage by Soranus (*Gyn.* I 27,3,4 = CMG IV 18,2 Ilberg: καθέδριον διάγειν βίον), the only author who uses the adjective καθέδριος with the meaning of "sedentary" (LSJ⁹ 851 *s.v.* καθέδριος 2), but it absolutely makes no sense in this context. This last case strongly illustrate how a good lexical awareness of the semantic values of a word in all its occurrences and kinds of sources can contribute to a better philological understanding of the text itself.

Philologically challenging and stimulating are the ll. 66–76 of P.Ryl. III 529r (col. ii):

οἱ] | μὲν ἄλλοι καθέ[δριον . . .]|ωσαν τὸν κάμ[νοντα] σχη||ματίζειν. ἡμεῖ[ς δὲ . . . . .] | κεκλιμένο[[.]] ‛ ν ’· τὸ Ἀλ[εξάν]|δριον σχῆμά ἐστιν δυ[σαλ]|γέστατον, τὸ δὲ κεκλ[ιμέ]|νον ἀσφαλέστερον. ἀ[σφα]|λέστερον δὲ ὁτὲ μὲ[ν ὕπτι]|ον σχηματίζειν τὸν [πάσ]|χοντα, ὁτὲ δὲ πρηνῇ.

> The other physicians put the patient in the sitting position, whereas we lay him/her down: the "Alexandrian position" is extremely painful, while the lying down position is safer. It is safer to place the patient sometimes on his/her back, sometimes on his/her stomach.

The Rylands papyrus was first published by ROBERTS 1938, 158–62 (*ed.pr.*), then republished by MARGANNE 1998 110–47 (*ed.alt.*) with some textual differences, as in the case of l. 67. Here the adjective καθέδριος was very plausibly restored by MARGANNE 1998, 112 and 117, whereas the *editor princeps* (ROBERTS 1938, 160) simply printed κατε[. The term is mentioned in the earliest extant discussion on the best and less painful position in which to place the patient before performing the reduction of his/her dislocated shoulder. Indeed, in Greek medical sources the sitting position, the καθέδριον σχῆμα, is sometimes explicitly rejected or preferred to other positions, especially the supine one.

A first observation on this passage: given the end ωσαν in l. 68 and the context, one could expect the first aorist active indicative of a verb of thinking/saying (e.g. "think", "claim", "advise") in the lacuna of l. 67, followed by the infinitive σχηματίζειν in ll. 68–9. Considering the previous presence of καθέδριον and what seems to be the space left at the end of the lacuna, which is measured with three dots in Marganne's

---

44 Cf. MARGANNE 1994, 106.

edition, a plausible supplement might be perhaps ἠξί‖ωσαν from ἀξιόω, with the meaning "to think", i.e. ἄλλοι καθέ[δριον ἠξί]‖ωσαν τὸν κάμ[νοντα] σχη‖ματίζειν. Interestingly, ROBERTS 1938, 16, though without any supplement to the text, translates: "other authorities (advise) that the patient be disposed in a ... position".

Moreover, at the end of l. 69 before the break, possible supplements might be: ἡμεῖ[ς δὲ μᾶλλον] | κεκλιμένο⟦.⟧`ν´ or ἡμεῖ[ς δὲ κατα]|κεκλιμένο⟦.⟧`ν´, cf. Aët. XV 5,50–1 (19,15–6 Kostomiris) σχηματιζέσθω ὁ πάσχων κατακεκλιμένος.

A more puzzling textual issue concerns ll. 70–1. The context (ll. 66–81) is the description of a method of reduction opposed to the procedures of other physicians or schools and it is specified that others place the patient seated; on the contrary, the author recommends lying him down. According to the restoration made in ll. 70–1 by both Roberts and Marganne, i.e. Ἀλ[εξάν]‖δριον, the so called Ἀλεξάνδριον σχῆμα is defined as δυσαλγέστατον, whereas the lying-down position is regarded as less painful and safer. Marganne (p. 129) considers the Ἀλεξάνδριον σχῆμα as the same as the sitting position mentioned in l. 67, and not as a different posture. In such a case, the second statement (τὸ Ἀλ[εξάν]‖δριον σχῆμά ἐστιν δυ[σαλ]‖γέστατον, τὸ δὲ κεκλ[ι-μέ]‖νον ἀσφαλέστερον) would seem to explain the reason for the author's preference for the lying-down position, as expressed in the previous lines (οἱ] | μὲν ἄλλοι καθέ-[δριον . . .]‖ωσαν τὸν κάμ[νοντα] σχη‖ματίζειν. ἡμεῖ[ς δὲ . . . . .] | κεκλιμένο⟦.⟧`ν´).

The point is that there is no attestation of the Ἀλεξάνδριον σχῆμα in medical literature, and Withington's suggestion[45] that the Ἀλεξάνδριον σχῆμα may correspond to the position on the Thessalian straightbacked chair, the μέγα ἕδος Θεσσαλικόν, used for dislocations and mentioned by Hippocrates (*Art*. 7,36–43 = IV 92,10–94,1 L., cf. also Gal. *In Hp. Art*. I 22 = XVIIIa 344,1–345,8 K.), seems unmotivated.[46]

The reading Ἀλ[εξάν]‖δριον, accepted by MARGANNE 1998, 112, was first restored by ROBERTS 1938, 160, who did not understand the presence of καθέ[δριον in l. 67, as aforesaid. But a collation with the digital image of the papyrus[47] has raised difficulties with the supplement Ἀλ[εξάν]‖δριον, inasmuch the traces are incompatible with the usual shape of α and λ. At the break, indeed, part of a small horizontal trace survives, consistent with the lower trait of an α. It is preceded by what seems to belong to a

---

45 Cf. ROBERTS 1938, 162.

46 Cf. MARGANNE 1998, 129 n. 29: "Le remarque de Withington [...] est superflue. Il est vrai que ni lui, ni Roberts n'avaient restitué, à la l. 67, καθέ[δριον" .

47 The only available image of this papyrus is black and white and is stored in the photographic archive of CEDOPAL (University of Liège). Under my request Prof. Marganne, to whom I express all my gratitude, provided me with the available image. Then, in order to have the best image possible to check the text, I purchased high-quality images of the recto and the verso of the Rylands papyrus (format: large TIFF 600 dpi) from the Centre for Heritage Imaging and Collection Care (CHICC) of the University of Manchester Library. I would like to take the opportunity to thank John Hodgson, Manuscript and Archives Manager, and Tony Richards, Heritage Photographer, for the excellent digital images that now I own.

vertical trait slightly sloping to right. The shape of these traces can be convincingly compared with the sequence γα recurring several times in the papyrus.[48] Accordingly, also taking into account the ending δριον in l. 71, a likely supplement that seems paleographically compatible and fits well the space is τὸ γὰ[ρ καθέ]|δριον. Additionally, on the syntactic side, the presence of γὰ[ρ would acquire an explanatory function perfectly appropriate to the content, thus introducing why the author – that in the previous sentence affirms his preference for the lying-down position – rejects the καθέδριον σχῆμα, being it δυσαλγέστατον, "more painful". This sense of opposition is then marked and strengthened by the adversative δέ (l. 72), where it is stated that the lying down position is safer.

To conclude on this aspect, the example of καθέδριος is useful to highlight the potential philological contribution of the vertically in-depth and both 'atomic' and 'holistic' lexical studies carried out for *Medicalia Online*. Thus, given the high degree of textual awareness offered by this kind of lexical studies, the present *specimen* illustrates how the medical text can be improved in terms of new and more reliable (or at least plausible) readings and emendations, as well as it can be deeper and better understood:

οἱ] | μὲν ἄλλοι καθέ[δριον ἠξί]|ωσαν τὸν κάμ[νοντα] σχη||ματίζειν. ἡμεῖ[ς δὲ μᾶλλον] | κεκλιμένο⟦ ˳ ⟧ ʽ ν ʼ· τὸ γὰ[ρ καθέ]|δριον σχῆμά ἐστιν δυ[σαλ]|γέστατον, τὸ δὲ κεκλ[ιμέ]|νον ἀσφαλέστερον.

The other physicians think that it is better to put the patient in the sitting position, but we rather that it is better to lay him/her down: the sitting position is indeed extremely painful, whereas the lying down position is safer.

3 Conclusions

The experience of a specialized lexicographical tool like *Medicalia Online* demonstrates the different kinds and levels of usefulness and the potential of a digital lexicographical database characterized by a wide-ranging approach in the era of Digital Humanities. Among these, the usefulness of a tool specifically conceived to carry out in-depth, 'vertical' lexical studies aimed at providing the broadest possible overview of the examined items under multiple perspectives, namely linguistic, archaeological, historical-scientific. The usefulness concealed in the inner nature of a digital instrument, that is its dynamism, its flexibility, its capability of allowing the author of the entry to constantly update the contents. The usefulness of the interdisciplinary,

---

48 Cf. *recto* col. I l. 5 γάρ, l. 15 ἀνάλογα, ll. 33 and 50 γάρ; *verso* col. I ll. 106 and 109 γάρ, as well as l. 110 ὀργανικοῖς, and finally col. II l. 124 τοιγαροῦν and l. 127 ἔργα.

comprehensive approach to the Ancient World, which makes *Medicalia Online* a helpful resource for contributing to improve and reconsider the studies in medical papyrology as well as in the other subject areas involved, and broadens the circle of users to any scholar or enthusiast of those research fields. Furthermore, the user-friendliness of the entries in spite of the wide range of information and, most importantly, the mutual enrichment resulted from the interconnection between the *MedOn* lemmas and the digital editions of the medical texts on papyrus in the core database. This strongly encourages an enhanced textual understanding stimulating a double awareness – a concrete one and a textual-philological one –, as the selected *specimina* have illustrated. The outcomes are, on the one hand, the attempt to reconstruct the physical reality hidden under the words and, on the other hand, a better and deeper exegesis of the ancient medical texts through a better and deeper knowledge of the words themselves.

Such a 'vertical' lexicographical approach also helps elucidating the degree of technicality of the medical terms, as well as of the texts containing them. Two of the presented case studies can be taken as illuminating examples of this aspect. καθέδριος (§ 2.3) has a marked technical value, since it occurs, with its specialized medical meaning, only in strictly medical sources; conversely, the compound φαρμακοθήκη (§ 2.2), characterized by a pronounced morphological transparency, did not developed into a genuine technical term and was probably employed as a colloquial noun for an object of medical use, the physician's tool-case or its household counterpart, so that it was not a medical *terminus technicus* but just a word with a medical meaning. And indeed it does not appear in a *stricto sensu* medical papyrus, but in a private letter carrying medical information, P.Oslo II 54. Horeion, the author of the letter, who lives away from home for personal or professional reasons, asks his father Apollonios to send him the portable medicine-chest (ll. 5–6 πέμψον | μοι τὴν φαρμακοθήκην), as well as two remedies with different properties, that is a biting *pharmakon* and a much milder one (ll. 7–9 αἰτήσας π[αρ]ὰ τοῦ ἰατροῦ | φάρμακον δακνηρὸν| καὶ ἕτερον ἡδύτερον). The terms employed by Horeion point to different degrees of technicality and particularly significant is the juxtaposition between the φάρμακον δακνηρόν and the ἡδύτερον one, likely representing two *kollyria* of contrasting virtues, perhaps belonging to the typology of the *acharista*.[49] These terms, indeed, especially the former, are technical adjectives applied to eyesalves and ophthalmic preparations. Thus, whilst the vocabulary employed by the writer reflects different levels of technicality, the choice of accurate and rare words reveals his proficiency in the medical field and a certain medical literacy.[50] This might mean that Horeion was either a literate layman interested in medicine, or even a person with healing skills, such as a would-be physician or a *pharmakopoles*. The case of P.Oslo II 54 is useful to demonstrate how a

---

49 Cf. Bᴏɴᴀᴛɪ 2016b, 667–9.
50 On the topic, cf. Hᴀɴsᴏɴ 2010.

thorough lexical study of medical papyri from Greco-Roman Egypt contributes to shed new light on the socio-linguistic context in which these papyri were inserted, and may provide invaluable glimpses of the individuals 'behind their words'.

Finally, all these factors restate again the usefulness of *Medicalia Online* and its methodology to plumb the depths of the past, allowing for the revaluation of the entire material come to the present and promoting a more integrated knowledge of the Ancient World, to which a voice – if not 'living' at least not 'dead', or irremediably lost –can be returned: the echo of its written and material soul.

4 Bibliography

ANDORLINI, I. (1996), *Il papiro di Strasburgo inv. G 90 e l'oftalmologia di Aezio*, in *Storia e ecdotica dei testi medici greci. Atti del II Convegno Internazionale (Parigi, 24–26 maggio 1994)*, a. c. di A. Garzya, Napoli, 7–30.

ANDORLINI, I. (1997a), c., *'Specimina' per il Corpus dei Papiri Greci di Medicina. Atti dell'Incontro di studio (Firenze, 28–29 marzo 1996)*, Firenze.

ANDORLINI, I. (1997b), *Progetto per il Corpus dei Papiri Greci di Medicina*, in *Akten des 21. Internationalen Papyrologenkongress (Berlin, 13. – 19.8.1995)*, hrsg. von B. Kramer, W. Luppe, H. Maehler und G. Poethke, Stuttgart – Leipzig, I, 17–24.

ANDORLINI, I. – REGGIANI, N. (2012), *Edizione e ricostruzione digitale dei testi papiracei*, in *Diritto romano e scienze antichistiche nell'era digitale. Convegno di studio (Firenze 12–13 settembre 2011)*, a. c. di N. Palazzolo, Torino, 131–46.

BERTONAZZI, F. (2018), *Digital Edition of P.Strasb. inv. 1187: Between the Papyrus and the Indirect Tradition*, in *Proceedings of the 28th International Congress of Papyrology (Barcelona, 1–6 August 2016)*, Barcelona, forthcoming.

BIBER, D. – CONRAD, S. – REPPEN, R. (1998), *Corpus Linguistics, Investigating Language Structure and Use*, Cambridge.

BLIQUEZ, L.W (1994), *Roman Surgical Instruments and Other Minor Objects in the National Archaeological Museum of Naples*, Mainz.

BONATI, I. (2016a), *Il lessico dei vasi e dei contenitori greci nei papiri.* Specimina *per un repertorio lessicale degli angionimi greci*, Berlin – Boston.

BONATI, I. (2016b), *Between Text and Context: P.Oslo II 54 Reconsidered*, in *Proceedings of the 27th International Congress of Papyrology*, ed. by T. Derda, A. Łajtar, and J. Urbanik, Warsaw, 659–75.

BONATI, I. (2018a), *Medicalia Online: tecnicismi medici tra passato e presente*, in *Greek Medical Papyri – Text, Context, Hypertext. Proceedings of the DIGMEDTEXT International Conference (Parma, 2–4 November 2016)*, ed. by N. Reggiani, Berlin – Boston, forthcoming.

BONATI, I. (2018b), *Medicalia Online: An Electronic Dictionary of Technical Terms in Medical Papyri*, in *Proceedings of the 28th International Congress of Papyrology (Barcelona, 1–6 August 2016)*, Barcelona, forthcoming.

BONATI, I. – MARAVELA, A. (2018), eds., *Medicalia. Studies of Medical Technical Words from the Greek Papyri to Modern Medicine*, Leiden – Boston, forthcoming.

CHANTRAINE, P. (1933), *La formation des noms en grec ancien*, Paris.

DELATTRE, A. – HEILPORN, P. (2014), *Electronic Resources for Graeco-Roman and Christian Egypt: A Review of the State of the Net (March 2014)*, BiOr 71, 308–31.

Del Corso, L. (2007), *Il progetto PSI On-Line: applicazioni informatiche per una filologia materiale dei testi papiracei*, in *Digital Philology and Medieval Texts. Proceedings of the International Seminar (Arezzo, 19–21 Gennaio 2006)*, ed. by A. Ciula and F. Stella, Pisa, 165–74.

Deneffe, V. (1893), *Étude sur la trousse d'un chirurgien gallo-romain du III siècle*, Anvers.

Depauw, M. – Gheldof, T. (2014), *Trismegistos. An Interdisciplinary Platform for Ancient World Texts and Related Information*, in *Theory and Practice of Digital Libraries. TPDL 2013 Selected Workshops*, ed by Ł. Bolikowski, V. Casarosa, P. Goodale, N. Houssos, P. Manghi, and J. Schirrwagen, Berlin, 40–52.

Evans, T.V. – Obbink, D.D. (2010), eds., *The Language of the Papyri*, Oxford.

Facchinetti, R. (2007), *Theoretical Description and Practical Applications of Linguistic Corpora*, Verona.

Fischer, K.D. (1997), *Was ist das δελτάριον in P.Oxy. LIX 4001?*, in Andorlini 1997a, 109–13.

Fischer Bovet, C. (2009), *Letter Requesting Medical Tools*, in *Greek Medical Papyri II*, ed. by I. Andorlini, Firenze, 157–65.

Fuertes-Olivera, P.A. (2013), *E-lexicography: The Continuing Challenge of Applying New Technology to Dictionary-Making*, in *The Bloomsbury Companion to Lexicography*, ed. by H. Jackson, London – New York, 323–40.

Fuertes-Olivera, P.A. – Bergenholtz, H. (2011), *E-Lexicography. The Internet, Digital Initiatives and Lexicography*, New York.

Fuertes-Olivera, P.A. – Tarp, S. (2014), *Theory and Practice of Specialised Online Dictionaries. Lexicography versus Terminography*, Berlin – Boston.

Hanson, A.E. (2010), *Doctors' Literacy and Papyri of Medical Content*, StudAncMed 35, 187–204

Klosa, A. (2013), *The Lexicographical Process (with special focus on online dictionaries)*, in *Dictionaries. An International Encyclopedia of Lexicography. Supplementary Volume*, ed. by R.H. Gouws, U. Heid, W. Schweickard, and H.E. Wiegand, Berlin – Boston, 517–24.

Kuebler, S. – Zinsmeister, H. (2014), *Corpus Linguistic and Linguistic Annotated Corpora*, London – New York.

Künzl, E. (1996), *Forschungsbericht zu den antiken medizinischen Instrumenten*, ANRW II 37.3, 2433–639.

Leith, D. (2009), *Medical and Related Texts*, in *The Oxyrhynchus Papyri. Vol. LXXIV*, London, 46–88.

Lemberg, I. (2001), *Aspekte der Online-Lexikographie für wissenschaftliche Wörterbücher*, in *Chancen und Perspektiven computergestützter Lexikographie*, hrsg. von I. Lemberg, B. Schröder und A. Storrer, Tübingen, 71–91.

Magnani, M. (2008), *"Sapere ex indicibus"*, in *Scienze umane e cultura digitale. Atti della XVI Settimana della Cultura Scientifica (Parma, 16–17 maggio 2006)*, a c. di A.M. Tammaro e S. Santoro, Fiesole – Firenze, 127–38.

Maravela, A. (2017), *Medical Micro-Language: The Contribution of Papyri to the Study of Medical Greek*, in *Parlare la medicina: fra lingue e culture, nello spazio e nel tempo. Atti del Convegno Internazionale (Parma, 5–7 Settembre 2016)*, a c. di N. Reggiani e F. Bertonazzi, Firenze, forthcoming.

Marganne, M.H. (1994), *L'ophtalmologie dans l'Égypte gréco-romaine d'après les papyrus littéraires grecs*, Leiden – New York – Köln.

Marganne, M.H. (1998), *La chirurgie dans l'Égypte gréco-romaine d'apres les papyrus littéraires grecs*, Leiden – Boston – Köln.

Marganne, M.H. (2004a), *Le médecin, la trousse et le livre dans le monde gréco-romain*, in *Da Ercolano all'Egitto IV. Ricerche varie di Papirologia* ["Papyrologica Lupiensia" 12], a c. di M. Capasso, Galatina, 117–30.

Marganne, M.H. (2004b), *Le livre médical dans le monde gréco-romain*, Liège.

McGann, J. (2010), *Introduction*, in *Online Humanities Scholarship: The Shape of Things to Come*, ed. by J. McGann, Houston, 1–4.

Milne, J.S. (1907), *Surgical Instruments in Greek and Roman Times*, Oxford.

NUTTON, V. (2004), *Ancient Medicine*, London – New York.

PRINSLOO, D.J. (2001), *The Compilation of Electronic Dictionaries for the African Languages*, "Lexikos. Journal of the African Association for Lexicography" 11, 139–59.

RAMSAY, S. (2004), *Databases*, in SCHREIBMAN – SIEMENS – UNSWORTH 2004, 177–97.

REGGIANI, N. (2012), *Le tavolette della mente. Risorse digitali e Antichistica: il caso della Papirologia*, in *Umanisti e risorse digitali*, a c. di A.M. Tammaro, Parma, 88–110.

REGGIANI, N. (2015), *A Corpus of Literary Papyri Online: the Pilot Project of the Medical Texts via SoSOL*, in *Antike Lebenswelten Althistorische und papyrologische Studien*, hrsg. von R. Lafer und K. Strobel, Berlin – New York, 341–52.

REGGIANI, N. (2016), *The Corpus of Greek Medical Papyri and Digital Papyrology: New Perspectives from an Ongoing Project*, in *Altertumswissenschaften in a Digital Age – Egyptology, Papyrology and Beyond. Proceedings of a Conference and Workshop in Leipzig, November 4–6, 2015*, ed. by M. Berti and F. Naether, Leipzig.

REGGIANI, N. (2017), *Digital Papyrology I. Methods, Tools and Trends*, Berlin – Boston.

REGGIANI, N. (2018a), *The Corpus of Greek Medical Papyri Online and the Digital Edition of Ancient Documents*, in *Proceedings of the 28th International Congress of Papyrology (Barcelona, 1–6 August 2016),* Barcelona, forthcoming.

REGGIANI, N. (2018b), *Digitizing Medical Papyri in Question-and-Answer Format*, in *"Where Does It Hurt?" Ancient Medicine in Questions and Answers*, ed. by E. Gielen and M. Meeusen, Leiden – Boston, forthcoming.

ROBERTS, C.H. (1938), *Catalogue of the Greek and Latin Papyri in the John Rylands Library*, III, Manchester.

SCHIRONI, F. (2010), *Technical Languages: Science and Medicine*, in *A Companion to the Ancient Greek Language*, ed. by E.J. Bakker, Oxford, 338–53.

SCHREIBMAN, S.– SIEMENS, R. – UNSWORTH, J. (2004), eds., *A Companion to Digital Humanities*, Oxford.

SCHRÖDER, M. (1997), *Brauchen wir ein neues Wörterbuchkartell? Zu den Perspektiven einer computerunterstützten Dialektlexikographie und eines Projektes „Deutsches Dialektwörterbuch"*, ZDL 64, 57–65.

STERN, E.M. (1999) *Ancient Glass in Athenian Temple Treasures*, "Journal of Glass Studies" 41, 19–50.

TARP, S. (2008), *Lexicography in the Borderland between Knowledge and Non-Knowledge. General Lexicographical Theory with Particular Focus on Learner's Lexicography*, Berlin – Boston.

TURNER, E.G. (1939), *Catalogue of Greek and Latin Papyri and Ostraca in the Possession of the University of Aberdeen*, Aberdeen.

SVENSSON, P. – GOLDBERG, D.T. (2015), eds., *Between Humanities and the Digital*, Cambridge (MA).

Indices

All along the volume, the standard bibliographical references to papyrus editions (after the online *Checklist of Editions* at http://papyri.info/docs/checklist + GMP for the *Greek Medical Papyri* series) are used. Other common abbreviations are those of the digital catalogues HGV = *Heidelberg Gesamtverzeichnis der Griechischen Papyrusurkunden Ägyptens* (http://aquila.zaw.uni-heidelberg.de/start); TM = *Trismegistos* (http://www.trismegistos.org); APIS = *Advanced Papyrological Information System* (via *Papyri.info*). Bibliographical abbreviations of academic journals follow the *Année Philologique*. LSJ stands for the ninth edition of the Liddell-Scott-Jones dictionary.

I. (Key)words and names

N.B. The most frequent terms (e.g. papyrus, digital, edition, editor, language, philology, etc.) are not indexed.

abbreviations 20, 24–5, 39, 51, 59, 79–80, 127–8, 132–3, 135, 143

Academia.edu 82, 90

accents 27, 35, 122

access, accessibility 9, 14, 53, 73, 79, 82, 93–4, 99, 101, 107, 151, 153

accretive model 12, 54

acute (accent) 35

additions 12, 43

adespota 150, 153

Aeneid 90

Aeschylus 96

Aëtius 35, 47

AGDT (*Ancient Greek and Latin Dependency Treebank*) v, 107–8, 140–1, 144–6, 157

aggregation 75

alignment v, 4, 10, 14, 17, 20, 24, 27, 36–7, 47, 61, 89

Almas, B. 157

alternatives 3–4, 26–7, 37, 43–6, 51, 53, 84, 121, 126, 144

Ammon (archive) 45

Ammonius 116

Anagnosis v, 20, 47, 63, 71–3, 80

Andorfer, P. 88

Andorlini, I. v–vi, 3, 10, 13, 15–16, 23, 64, 95, 149–50

angle/angular bracket 24, 31, 46–7, 121, 124–5, 134

annotation (digital) v–vi, 3, 14, 19–24, 29–30, 32, 36–7, 41, 45, 51, 53, 58, 61, 73, 89, 92, 105–14, 116–8, 120, 123, 135–6, 139–42, 146–7, 157–8, 172

annotations (ancient) 11–12, 33, 41–3, 48, 52, 59, 61, 73, 101, 105–9, 116, 118, 120, 136, 140–1, 157, 172

Anonymus Londinensis 43, 45

Antinoupolis 24, 41, 57

Antyllus 38, 165

aorist 111, 116, 125, 167

aphorisms (Hippocrates) 28, 41

Apion (archive) 126

APIS (*Advanced Papyrological Information System*) 76

Apollonius 116

Apollonius Mys 21, 38

apostrophe 30

apparatus criticus v, 3–4, 7, 10, 26–7, 29, 32, 35, 37, 45, 57, 67–8, 77–80, 87, 89–90, 92–4, 96–8, 109, 116, 119–26, 133–5, 166

appearance 17, 20, 33–4, 54

Arabic Papyrology School 80

Arabic texts 37, 80

archaeology 152–3, 159–60, 162, 169, 171

archetype 3, 5, 7, 10, 37, 53

archives 45, 60, 101, 112, 118, 126, 136, 137, 168

Arethusa 108–10, 142

AristarchusX 57, 71

artefacts 9, 53, 59, 95

artificer (of facts, of data) 13, 61, 137

asper (spiritus) 35, 55

II. Papyrus texts

The papyri are cited according to the official *Checklist of Editions* (http://papyri.info/ docs/checklist), to which add GMP = *Greek Medical Papyri I–II*, ed. I. Andorlini, Firenze 2001–2009. The list has been compiled with the precious help of Alessia Bovo, for which I am most grateful.

III. Literary sources

IV. Greek concordance

N.B. Single Greek letters are not indexed.

V. XML tags

```
<ab>  18
<abbr>  24–6
<add>  42–3
<add place="above">  42
<add place="below">  42
<add place="inline">  44
<add place="interlinear">  42
<add place="left">  42
<add place="right">  42
<am>  25–6
<app>  37, 45
<app type="variant">  37
<bibl>  144
<choice>  26–7, 31–2, 143
<corr>  26, 31, 121
<d n="xx" v="xx">  145
<del>  44
<del rend="corrected">  44
```

```
<del rend="cross-strokes">  44
<del rend="erasure">  44
<del rend="slashes">  44
<desc>  19, 27
<div type="textpart">  18
<e>  145
<ex>  24–6, 31
<expan>  24–6, 31
<f>  145
<figDesc>  35
<figure>  35
<foreing xml:lang="">  27
<g>  30, 36
<g type="check"/>  31
<g type="dipunct"/>  30
<g type="horizontal-stroke"/>  25
<g type="middot"/>  30
<g type="mid-punct">  142
```