Encoding linguistic variation in Greek documentary Papyri: the past, present and future of editorial regularization

Stolk, J.V.; Reggiani, N.

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, λειβα to λιβα and [ἀπ]οδόσω to [ἀπ]οδώσω in SB XXVI 16796,10–11, 16, are regularly found among these

---

20 All editorial mistakes and problematic instances marked out in this article can of course be revised through the *Papyrological Editor*, reducing the amount of variation slightly. These examples are, however, understood to be representative for some more fundamental problems with the practice of linguistic regularization. These problems and their possible solutions will be discussed further in section 5.
21 http://www.trismegistos.org/textirregularities, state of PN January 2014. Part of the search queries for this paper are made in the offline database, state May 2017.
22 For some of these texts someone from the editorial board already suggested changing the correction tags into regularization tags before finalization of the entry, see for example the editorial history of O.Did. 417 and P.Naqlun II 22, but these changes did not find their way into the online edition.

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 ἑκατασταῖς was regularized as "*l.* ναύλοις καὶ ἑκατοστaῖς" in the apparatus of P.Jena II 8,7, but ναύλοις was entered into the PN as a correction, καὶ as regularization and ἑκατοστῆς 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 ἀπλῆ 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 o 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 justi-fied 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 at-tempts 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 centu-ries 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 regular-ization 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 edi-tions, 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). Regu-larizations 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 ἰς, 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.