A cognitive approach to spelling production in historical sources: explaining the variation between and in Greek documentary papyri

Stolk, J.V.

Transactions of the Philological Society Volume 119:3 (2021) 289–314 doi: 10.1111/1467-968X.12219

A COGNITIVE APPROACH TO SPELLING PRODUCTION IN HISTORICAL SOURCES: EXPLAINING THE VARIATION BETWEEN <ε, αι> AND <ο, ω> IN GREEK DOCUMENTARY PAPYRI

By Joanne Vera Stolk
Ghent University

(Submitted: 27 October, 2020; Accepted: 23 June, 2021)

Abstract

The Greek documentary papyri (300 BCE – 700 CE) provide an interesting corpus for linguistic study due to the large amount of linguistic variation. Variation in spelling is traditionally used as evidence for phonological changes in the post-Classical Greek language. The interchange of graphemes, however, does not only depend on the pronunciation of the corresponding phoneme. In this paper I examine the cognitive processes behind the production of non-standard Greek orthography in more detail by applying an interactive dual-route model for spelling. If two graphemes are pronounced identically in the spoken language, the final choice between one grapheme or the other is likely to be based on cognitive and social aspects, such as the general frequency and probability of spelling patterns in the language, previous exposure of the writer to other lexemes and morphemes in the language and local scribal conventions, to name just a few factors. On the basis of examples of the frequent interchanges of <*e*, *ai*> and <*o*, *ō*>, I show how these other factors can contribute to a better interpretation of spelling production in documentary papyri.

1. Introduction

Greek documentary papyri are well-known for the presence of linguistic variation and variant spellings have been very helpful for reconstruction of phonological changes in the spoken language (see e.g. Mayser & Schmoll 1970, Gignac 1976 and Teodorsson 1977). By focusing explicitly on the reconstruction of the phonological system in the spoken language, however, these studies generally fail to explain important aspects of the production of spelling in the written language. Even when two phonemes have merged completely in pronunciation, the writer still has two (and sometimes even more) graphemes to choose from. How does the writer select the one or the other grapheme to represent a particular sound?

In this article, I will try to reveal other mechanisms that may have an impact on the variation in spelling in documentary papyri, besides phonological change. Since many Greek morphemes are distinguished by only one or two graphemes, linguistic research on changes in morphology or morphosyntax also depends on the exact spelling of the forms that are being studied. For example, the interchanges of the graphemes <ο, ω, ου> *o*, *ō*, *ou* may impact our understanding of the use of cases in the second declension, while interchanges of the graphemes <ε, η> *e*, *ē* and <ο, ω> *o*, *ō* may also affect our interpretation of the use of indicative and subjunctive. The variation between the graphemes <ο, ω> *o*, *ō* and <ε, αι> *e*, *ai* will be taken as case studies here. Apart from the variation between <ι, ει> *i*, *ei*, these are the two most frequent interchanges in the papyrus corpus (Gignac 1976: 192; Depauw & Stolk

2015: 207) and, interestingly, they also affect both nominal and verbal inflection. The interchanges are treated by Gignac very briefly, concluding the following:

> There is a very frequent interchange of αι and ε in all phonetic environments from the beginning of the Roman period on, indicating the identification of the classical short diphthong /ai/ with the simple vowel /ε/. (Gignac 1976: 191–3)

The same is observed for the interchange of <ο> and <ω> (Gignac 1976: 275–7). While the distinction between /ε/ and the /ai/ may have been preserved until the end of the Ptolemaic period (Mayser & Schmoll 1970: 85–7), or at least the later second century BCE (Teodorsson 1977: 222–5), the loss of quantitative distinction of /o:/ and the merger with /o/ probably already started during the third century BCE and examples start to accumulate during the second century BCE, especially in unaccented syllables (Mayser & Schmoll 1970: 73–6; Teodorsson 1977: 233–4). These interchanges are not just found in the papyri in Egypt, but they occur also in other historical sources and the results of the sound mergers underlying them are still present in spoken Modern Greek.

Although these spelling interchanges are among the most frequent ones attested in the papyri, amounting to over 20,000 instances, the production of either one of these graphemes <ο> or <ω> and <ε> or <αι> in any text is generally poorly understood. Even if <ο> and <ω> are likely to have been pronounced exactly the same in all phonetic contexts, this does not help us explain why the following phrase would be spelt like in example (1).[1]

(1)	ἐπομνύμενος	θεὸν	τὸν	παντωκράτωρα	καὶ	τὴν	εὐσέβιαν
	epomnúmenos	*theo͂n*	*to͂n*	*pantōkrátōra*	*kaì*	*te͂n*	*eusébian*
reg.		*theòn*	*tòn*	*pantokrátora*			*eusébeian*
	swearing	god-	the-	almighty-	and	the-	piety-
		ACC.SG.M	ACC.SG.M	ACC.SG.M		ACC.SG.F	ACC.SG.F

'swearing by God the Almighty and the piety' (P.Oxy. XVI 1880, 13; 427 CE)[2]

Why is the <ο> written twice in ἐπομνύμενος *epomnúmenos*, but replaced by <ω> in παντωκράτωρα *pantōkrátōra*? Why does the accusative singular look like genitive plural in θεὸν τὸν *theo͂n to͂n*, while we still have a clear accusative singular in the spelling of the next element τὴν εὐσέβιαν *te͂n eusébian*? Even though Gignac (1976: 277) noted this papyrus for its prominent variation between <ο, ω>, this does not answer the question whether the production of <ω> or <ο> in these types of texts is a random result of the identical pronunciation of the sound or whether we could identify some general principles behind the choice of graphemes in specific contexts.

Bülow-Jacobsen (2001) compared the spelling in respectively 31 and 56 short letters written on ostraca by the writers Philokles and Ischyras to nine letters on papyrus written by Gemellus. One of the patterns that he observed is that while Philokles often writes <ε> instead of <αι> and Ischyras <ο> instead of <ω>, Gemellus seems to prefer the opposite direction,

[1] The text is reproduced as on the papyrus. The line indicated with 'reg.' adds the regularized spelling of the words.

[2] The Greek text is taken from the Papyrological Navigator (PN; www.papyri.info) and the readings are checked on a photo and/or in the *editio princeps* and the *Berichtigungsliste der griechischen Papyrusurkunden aus Ägypten*. Text between square brackets is not preserved on the papyrus but supplemented by the editor and a dot under a letter signifies an uncertain reading. Accents are added in the transcription and transliteration, but they were not written on the papyrus. Length distinction in vowels was written, but probably not pronounced anymore. Transliteration, basic glosses and translation are provided for whole words; translations are my own, but may be based on the edition. Papyrus editions are cited according to the *Checklist of Greek, Latin, Demotic and Coptic Papyri, Ostraca and Tablets* www.papyri.info/docs/checklist [accessed 1 September 2020]; metadata are based on the information available in *Trismegistos* and the *Heidelberger Gesamtverzeichnis der griechischen Papyrusurkunden Ägyptens* (both accessible through the PN).

writing <αι> instead of <ε> and <ω> instead of <ο> more than the other two. Bülow-Jacobsen (2001: 161) concludes about the writers of the ostraca:

> they are not normally subject to pretentious hypercorrection, so that they rarely introduce a false ει where a simple ι suffices. The same goes for αι, so frequently written as they must have pronounced it, ε, but hardly ever the other way around. But Bellenus [Gemellus, JVS] is different, perhaps he was pretentious, perhaps his linguistic background was different.

Even if Gemellus would sometimes make different spelling choices than others, does this automatically mean that he pronounced things differently or had a different linguistic background? The analyses by Bülow-Jacobsen (2001) and Horrocks (2010: 172–3) hardly move beyond the phonological aspects of Gemellus' spelling, even though Bülow-Jacobsen suggests that those types of individual differences in spelling are more likely to be the result of spelling habits than pronunciation.

Scribes are human beings and we may not be able to explain every aspect of their written output. Still, there is no reason to assume that the cognitive mechanisms of the scribes in antiquity were radically different than the ones observed by cognitive linguists and psycholinguists nowadays. General patterns of frequency and analogy can be identified throughout the papyrus corpus and they may help us to better understand the production of spelling variants and their spread across linguistic and social contexts. I will here add insights from psycholinguistic theory to the sociolinguistic approach more generally applied to linguistic variation in the papyri in order to give a more complete usage-based account of the variation. The theoretical background of this approach will be discussed in Section 2 and the corpus of this study is introduced in Section 3. Section 4 will give an overview of the frequency of the interchanges in various linguistic contexts, before treating some of the most interesting aspects in more detail in the following sections about specific lexemes (Section 5), morphemes (Section 6) and the spread of alternative orthographic conventions (Section 7). Section 8 will apply these new results to the case of Gemellus presented above, before drawing more general conclusions in Section 9.

2. Understanding spelling production

Historical documents, like the Greek documentary papyri from Egypt, provide an excellent resource for the study of historical written language. Spelling, i.e., the graphic realisation of sounds, is an inextricable aspect of the written language. Variation in spelling may reflect phonological variation in the spoken language at the time, but it is now becoming increasingly accepted that historical written sources provide first and foremost evidence for the study of patterns of variation in the written language (Stenroos 2018). Spelling variation may thus be governed by its own social, extra-linguistic factors, such as the local conventions within a community of writers or the process of codification of a standard language.

While the post-Classical Greek language should not be regarded as a fully codified 'modern' standard language (Clackson 2015), the presence of strong social conventions with a large geographic and chronological distribution cannot be ignored. While some of the variation in the papyri could be explained by (temporary) changes in these orthographic conventions (Stolk 2020; see also Section 7 below), a large part of the identified 'non-standard' spellings in the papyri does not seem to attest to any coherent changes in contemporary orthographic norms. Previous study of social variables governing these types of variants invariably seems to lead to (in)competence, such as individual differences in knowledge about the norms of the written language (Evans 2012) or a general lower level of education acceptable for professional scribes (Bucking 2007). Although important, pointing

at a lower level of education at the level of the individual or the society does not automatically provide a satisfactory explanation of the linguistic output in question.

The written language and its orthography have to be learned and reproduced by the writer. Variation in the output is therefore not only the result of social factors, but is also determined by cognitive processes during language acquisition and language production. Following the uniformitarian principle, cognitive models and mechanisms identified in written language production in more recent times should also be applicable to the study of the production in historical written sources (Mair 2017). For example, the different processes during the production of spelling variants in papyri can be better understood with the help of a dual-route model for spelling developed in the field of cognitive neuropsychology (Section 2.1). Psycholinguistic research on spelling errors in Modern Greek also provides a useful comparison for the more language-specific aspects of Greek phoneme-to-grapheme correspondences. Common cognitive processing principles, such as frequency (Section 2.2) and analogy (Section 2.3), can also be applied to historical spelling variation. The combination of insights from both psycholinguistics and sociolinguistics not only allows for a more comprehensive usage-based theory of language change (Hundt et al. 2017), but, as I will show in this paper, it also enables a better understanding of the variation encountered in historical written languages.

2.1. Dual-route model

Spelling could be produced using two different routes or methods, each relying on a different source of orthographic knowledge. This two-level explanatory framework, also called the 'dual-route model for spelling', is widely accepted (see e.g. Barry 1994, Tainturier & Rapp 2001, Grainger & Ziegler 2011). The first 'lexical' orthographic method relies on a mental representation of the spelling of known words and morphemes. After a writer has been exposed to the orthography of a specific (part of a) word and it has been stored in orthographic long-term memory, it could be retrieved directly from this orthographic lexicon during language production. The second, 'sublexical' or 'non-lexical', orthographic process relies on the writer's knowledge of general spelling patterns in the language, such as letter-sound correspondences, letter combinations, and the probabilities of these patterns (Apel et al. 2019). The sublexical knowledge allows a writer to assemble the spelling of unfamiliar words or even non-existent words. How difficult it is to produce standard spellings by using the sublexical route only, depends on the language.

2.2. Frequency and pattern probability

Frequency effects are important to understand language processing and language use (see e.g. Bybee 2010). Psycholinguistic research shows that humans are sensitive to the frequency of occurrence of linguistic units and they implicitly learn the statistical patterns of the language they are exposed to, such as orthographic sequential probabilities (Ellis 2017). Frequencies are often our best approach to study both linguistic exposure and language production by writers in historical times (Hilpert 2017). For example, word frequency affects the likelihood that the spelling of a word is present in the orthographic lexicon of the scribe.

Frequency also plays a role in sublexical spelling production, although the relevant factors for sublexical orthographic knowledge tend to be more language specific. Languages differ in orthographic transparency, both with respect to the consistency of mapping graphemes onto phonemes (reading) as well as the mapping of phonemes onto graphemes (spelling). Modern Greek is characterised by a consistent mapping of graphemes to phonemes, like Finnish, for example, but an inconsistent mapping of phonemes to graphemes, similar to English

(Georgiou et al. 2012). For example, the phonemes /ε/ and /ɔ/ can be expressed by the graphemes <ε> or <αι> and <ο> or <ω>, respectively, in Modern Greek. Even though there are two options in each of these cases, choosing the right spelling does not need to be entirely random. The proportional distribution of these phoneme-to-grapheme mappings amounts to 78 vs. 22 per cent for <ε> vs. <αι> and 74 vs. 26 per cent for the distribution of <ο> vs. <ω> in Modern Greek (Protopapas & Vlahou 2009: 997). The correspondence that is most frequent, i.e. <ε> for /ε/ and <ο> for /ɔ/, is called the high contingency spelling pattern, while other, less frequent correspondences, are called low contingency patterns (Perry et al. 2002). Word frequency together with sound-to-spelling contingency are important predictors for error rates (Bonin et al. 2013). Words with low contingency spellings are slower to produce than words with high contingency spellings and this effect is stronger for words with a low frequency (Delattre et al. 2006). High contingency spelling patterns are also more often used to spell non-existent words, and thus when using the sublexical spelling method (Barry & Seymour 1988). The presence of multiple low contingency spelling patterns in Modern Greek, especially for vowels, leads to a high rate of (phonologically acceptable) misspellings (Protopapas et al. 2013). Since the main phonological mergers leading up to this situation of sound-spelling inconsistency in Modern Greek are assumed to have already been established in post-Classical Greek, similar cognitive and linguistic factors can be expected to be relevant to understand the production of variant spellings in the papyri, especially during the Roman and Byzantine period (1st–7th centuries CE).

Frequency effects can go beyond word frequency and the proportional distribution of graphemes. Skilled spellers also use more detailed knowledge, for example of the position in the word or the morphological status of the linguistic unit, in order to produce infrequent phoneme-to-grapheme correspondences (Perry et al. 2002). Corpus studies have shown that spelling errors are not only more likely to occur in low frequency words, but also in the less predictable roots than in the repetitive patterns of morphemes (e.g. Bar-On & Kuperman 2019). During spelling acquisition, Greek children first seem to adopt high contingency spellings for phonemes, e.g. <ι> and <η> for /i/, but morphological awareness helps them to perform better in spelling low contingency graphemes in morpheme positions than in word stems (Chliounaki & Bryant 2003). For more advanced spellers of Greek, the word-particular (idiosyncratic) spellings in roots tend to become easier along with increased lexical knowledge, while the spelling patterns found in inflectional – and especially derivational – suffixes take a longer time to master completely (Protopapas et al. 2013). Although this process of spelling acquisition could obviously be influenced by explicit educational methods, which may have been different in antiquity, the important roles of frequency and pattern probability are likely to result from implicit learning primarily and may thus also play a role in historical languages. General knowledge of the spelling patterns in the language is also used by adult second language learners (Yin et al. 2020). Thus, although the pronunciation of Greek speakers in Egypt may have been influenced by their Egyptian native language, their spelling choices are still likely to reflect (amongst others) the probability of spelling patterns in the Greek language they have been exposed to. This hypothesis will be tested in Section 4.

2.3. Analogy and priming

Although the two levels of orthographic knowledge, i.e. lexical and non-lexical, can function separately, they can also interact (Barry 1994: 40–4; Barry & Seymour 1988). For example, specific spelling patterns of words in the orthographic lexicon can be used to produce less familiar lexemes by analogy (Bosse et al. 2003). Analogy can be defined as a mapping of correspondences between a source and a target (Gentner 1983). This mapping involves a process of comparison of the two structures and the identification of perceptual or relational

similarities (Behrens 2017). Although there may be many possibilities for analogy, not all possible comparisons are always made and applied. Mappings with a superficial perceptual similarity seem to be more easily made (Behrens 2017). Lexical analogy is found more commonly when the source for analogy is lexically more frequent and/or phonetically more similar to the target (Tainturier et al. 2013). It is generally more likely to apply to low frequency items than high frequency items ('the conserving effect' of high frequency; Bybee 2010: 75). The effect of analogy is indeed found to be stronger for low contingency spellings than for high contingency spellings ('the inverse frequency effect'; Martin & Barry 2012), which makes analogy an interesting explanation for the (sublexical) production of low frequency spellings. General positional or morphological awareness (see 2.2) could be used to identify relational similarities between (parts of) words and produce morphologically plausible spellings (see Section 6). Properties such as word class, positional or morphological information, however, do not need to be taken into account by the writer, leading, for example, to the production of phonologically acceptable but non-existent case and verb endings.

The source for analogy could be present in the orthographic lexicon of the writer, but it could also be (made) available in the preceding linguistic context. The reproduction of a structure shortly after exposure to the same structure is called a priming effect (Ellis 2017: 119). This also means that (repeated) exposure to a variant spelling – or a spelling mistake – is likely to influence our mental orthographical representations (Dixon & Kaminska 1997) and can thus lead to the spread of (local) orthographic changes in individuals, despite previous exposure or education (see Section 7). Priming is often used to measure the effect of orthographical and morphological analogy in linguistic processing (e.g. Hasenäcker et al. 2016), but analogical effects are also found without explicit priming (e.g. Bosse et al. 2003; Tainturier et al. 2013). Working with historical written corpora does not allow the researcher to study priming in a controlled environment, since the situation of a spontaneous presentation of a spelling variant preceded by its potential analogical source is relatively rare, though not completely impossible (see a few possible examples in Sections 5 and 8). The potential of general lexical and morphological analogy to explain spelling production is examined further in Sections 5 and 6.

3. Corpus and methodology

Despite the application of quantifiable, replicable analytical methods, the study of cognitive mechanisms based on corpus data remains an indirect method of linguistic interpretation that brings several problems and challenges (see e.g. Arppe et al. 2010), especially when forced to work with the particularities of a historical corpus. Like many other corpora, the corpus of documentary papyri is not a perfectly represented randomised corpus (see e.g. Worp 2013). The prevalence of certain genres and formulaic phrases means that certain lexemes, and thus strings of graphemes, can be overrepresented in (parts of) the corpus. The occurrences of spelling variation are not randomly spread across the available linguistic contexts either. This tends to be obscured in the phonological descriptions of the corpus by, for example, Mayser & Schmoll (1970) and Gignac (1976), because they have manually selected a more representative cross-section of examples from the corpus, which may give the appearance that these variant spellings are randomly spread across the linguistic contexts to which the sound change applies. The frequencies that can be measured in a written corpus are only a partial representation of language use in general and of the probability of occurrence that the individual writers would have experienced in daily life (Hilpert 2017: 53–4). On the other hand, written corpora present us with real samples of language use in daily life and may give a more realistic reflection of the language a particular user would be exposed to than isolated

instances produced in an idealised context (Bresnan 2007). And thus frequency and probability measurements in a specific corpus may give more explanatory insight into language production than the external and judgmental 'standard' vs. 'non-standard' categorisation traditionally applied in orthographic studies.

This study is based on the full corpus of documentary papyri as digitalised in the Duke Databank of Documentary Papyri (DDbDP), accessible through www.papyri.info. Most texts are dated between the third century BCE and the eighth century CE and were found (and written) in Egypt. This corpus has been analysed within Trismegistos (TM) with the help of a database collecting editorial regularisations of spelling and morphology (TM Text Irregularities) and a lemmatised database with morphological annotation (TM Words).[3] Apart from the possible overrepresentation of particular lexemes and the non-random spread of spelling variation across lexemes in the corpus, the editorial regularisations of spelling are not entirely consistent either. There has not been one generally formulated norm for the judgement of the 'standardness' of spelling variants, which has resulted in some degree of variation in the annotation of 'non-standard' spellings (Stolk 2018). The most common results of this lack of editorial norms seem to have been the underestimation of the number of occurrences of very frequently occurring variants (remedied to some extent in the digital corpus, see Stolk 2018) and the inconsistent regularisation of lexemes for which the standard spelling is not clear in Greek or may have changed over time (Stolk 2020). This means that the accuracy of the editorial corpus data for spelling variation ultimately depends on the frequency of the lexeme and how frequently it is spelt differently. With the help of the lexically and morphologically annotated TM Words database, the groups of interchanges in the linguistic contexts introduced in Section 4 have all been cross-checked for the overrepresentation of certain lexemes or morphemes in these groups and the results of this secondary analysis form an essential part of this study of spelling variation (see Sections 5–7).

4. Frequency effects

Leaving the scribe's lexical knowledge and contemporary spelling conventions aside for now, we will start by focusing on the sublexical spelling process trying to establish the influence of this type of knowledge about general patterns in the Greek language on spelling production in Greek papyri. The graphemes <ε> and <αι> and <ο> and <ω> can be used to write the phonemes /ε/ and /o/ in Greek. If a scribe cannot retrieve the spelling of a lexeme from the orthographic lexicon, how can he/she select one of these graphemes to write down the sound? The hypothesis for this section is that the frequency distribution of the two graphemes is an important factor in this selection process. At a general level, this hypothesis is easily confirmed by looking at the frequency distribution of the graphemes in question.

The grapheme <ε> is used in approximately 73 per cent and <αι> in 27 per cent of the total occurrences of this sound in the Greek language of the papyri.[4] This means that <ε> would be the high contingency spelling and <αι> the low contingency spelling of the represented phoneme. Based on this distribution, we would thus expect writers to choose <ε> instead of <αι> more often than they choose to write <αι> instead of <ε>. This is indeed the case: there

[3] See www.trismegistos.org/textirregularities/ and www.trismegistos.org/words/, both databases were based on a scrape of all digitalized documentary papyri with transcription in the PN (state 2016) with a total number of words of 4,885,874.

[4] The total number of attestations is an approximation based on the sum of the number of words with <ε> or <αι> occurring in a particular context, see Table 1. A similar distribution of 74 per cent vs. 26 per cent, however, was calculated for the total occurrence of <ο> and <ω> in the (contemporary written) four gospels of the New Testament (see Stolk et al. in preparation). In Modern Greek the proportional difference is 78 per cent vs. 22 per cent (Protopapas & Vlahou 2009).

are 4,742 editorial regularisations of <ε> instead of <αι> and only 2,313 regularisations of <αι> instead of <ε> in papyrus editions. Table 1 shows the percentages for both interchanges of <ε> and <αι>.

Table 1. Frequency of interchanges compared to non-interchanged graphemes <ε, αι>.

Expected	<αι>		<ε>	
Written	<ε>	<αι>	<αι>	<ε>
N	4,742	457,395	2,313	1,272,731
%	1.03	98.97	0.18	99.82

A similar effect can be observed for the variant spellings of <o, ω>, although the distribution is more equal in this case. The grapheme <o> is the high contingency spelling used in approximately 62 per cent and <ω> the low contingency spelling used in 38 per cent of the total occurrences of this sound in the Greek language of the papyri.[5] We would thus expect writers to choose <o> instead of <ω> more often than <ω> instead of <o>. There are indeed 8,236 editorial regularisations of <o> instead of <ω> and only 6,632 regularisations of <ω> instead of <o> in papyrus editions. Table 2 shows the percentages for both interchanges of <o> and <ω>.

Table 2. Frequency of interchanges compared to non-interchanged graphemes <o, ω>.

Expected	<ω>		<o>	
Written	<o>	<ω>	<ω>	<o>
N	8,236	807,994	6,632	1,343,653
%	1.01	98.99	0.49	99.51

First of all, these comparisons show that the interchanges of the variant spellings <ε, αι> and <o, ω>, even though they may seem frequent compared to other spelling variants, are still a very small minority compared to the enormous number of cases where the expected graphemes are used in papyri. Only about 0.2–1 per cent of the attestations show variation. Even though the differences may seem small, the results are not random. The figure of 1.01 per cent of the potential attestations of <ω> written as <o> is significantly more than the 0.49 per cent of the instances of where <ω> is written for <o> and the same applies to the distribution of the interchange of <ε, αι>.[6]

This difference in the general frequency of the two graphemes could have been used by the writers in the process of written language production. The proportional frequency of two corresponding graphemes, however, may also vary with the precise linguistic context and writers could have applied this more specific knowledge of typical patterns for grapheme selection as well. The frequencies of occurrence and the interchanges in various linguistic contexts are compared for <ε, αι> in Section 4.1 and <o, ω> in Section 4.2.

[5] The total number of attestations is an approximation based on the sum of the number of words with <o> or <ω> occurring in a particular context, see Table 4 below. A similar distribution of 65 per cent vs. 35 per cent, however, was calculated for the total occurrence of <o> and <ω> in the (contemporary written) four gospels of the New Testament (see Stolk et al. in preparation). In Modern Greek the proportional difference is more pronounced with 74 per cent vs. 26 per cent (Protopapas & Vlahou 2009).

[6] The p-values for both are < 2.2e-16 in a Fisher's Exact test.

4.1. Frequency of interchanges of <ε, αι>

Table 3 shows the frequency of interchange for <ε, αι> in different linguistic contexts. As we can see, the interchanges occur in many different contexts, so the monophthongisation and merger of the two phonemes do not seem to be conditioned by a specific linguistic context only, as was already concluded by previous studies (see introduction). In all contexts, the difference between the two types of interchanges is significant ($p < 0.05$). In the vast majority of the cases, <ε> is more common than <αι> and also the interchange of <ε> instead of <αι> is (much) more frequent than the other way around, just as we observed above for these interchanges in general. Even in two contexts where the total number of attestations of <αι> is in fact slightly higher than <ε>, namely after <β> and after <χ>, the interchange of <ε> for <αι> is still more common than the reverse. In case of doubt, a writer could have chosen to write <ε> because this grapheme is more common generally.

Specific contexts can have a frequency effect that is weaker, but also demonstrably stronger than the general effect of the high contingency spelling on grapheme selection. For example, the prefix πραι- occurs predominantly in Latin loanwords in our corpus, such as πραιπόσιτος *praepositus* (military title), πραιτώριον *praetorium* (residence of a governor) or πραισίδιον

Table 3. Frequency of interchanges of <ε, αι> in different linguistic contexts.

Expected	<αι>			<ε>			
Written	N <ε>	% <ε>	N <αι>	N <αι>	% <αι>	N <ε>	p-value
Total	4,742	1.03	457,395	2,313	0.18	1,272,731	< 2.2e-16
Word-initial	230	1.96	11,493	312	0.08	373,388	< 2.2e-16
after π	244	3.01	7,851	83	0.07	115,776	< 2.2e-16
after τ	511	0.83	60,981	392	0.39	101,254	< 2.2e-16
after δ	47	1.83	2,526	93	0.12	77,134	< 2.2e-16
after θ	189	1.19	15,741	129	0.26	50,122	< 2.2e-16
after σ	318	3.38	9,084	307	0.47	64,629	< 2.2e-16
after ψ	93	3.33	2,699	12	0.12	10,162	< 2.2e-16
after ξ	100	6.63	1,408	15	0.17	8,624	< 2.2e-16
after ρ	231	1.41	16,182	73	0.21	33,968	< 2.2e-16
after λ	598	5.69	9,910	37	0.08	48,601	< 2.2e-16
after μ	614	0.92	66,034	268	0.15	172,939	< 2.2e-16
after ν	276	1.28	21,290	99	0.20	49,338	< 2.2e-16
after -ι	166	2.69	6,011	59	0.24	24,200	< 2.2e-16
after β	183	0.79	22,901	65	0.32	20,492	< 2.2e-16
after χ	120	0.95	12,555	25	0.33	7,614	1.018e-07
after κ	541	0.29	186,385	194	0.42	45,639	9.684e-06
Word-final	2,248	0.63	352,018	987	1.03	95,160	< 2.2e-16
before τ	156	3.20	4,722	90	0.06	144,775	< 2.2e-16
before δ	138	1.88	7,209	55	0.43	12,810	< 2.2e-16
before κ	61	2.28	2,610	66	0.12	53,616	< 2.2e-16
before σ	163	0.67	24,344	97	0.10	92,579	< 2.2e-16
before ρ	418	2.48	16,424	111	0.06	172,985	< 2.2e-16
before λ	15	0.85	1,746	93	0.13	69,702	7.122e-08
before μ	31	1.41	2,163	63	0.12	52,604	< 2.2e-16
before ν	249	5.39	4,368	143	0.07	218,373	< 2.2e-16
before α-	116	1.78	6,401	87	0.49	17,696	< 2.2e-16
before ο-	454	2.04	21,815	56	0.24	23,112	< 2.2e-16
before ω	230	2.91	7,672	63	0.08	80,371	< 2.2e-16

Note: Linguistic contexts with less than 60 interchanges for both directions are left out. Low numbers of interchanges are especially common for those contexts in which (one of) the graphemes is also infrequently attested. Furthermore, these smaller groups are easily dominated by single lexemes or a large number of occurrences in one particular social context (or even in a single text or by the same writer), see also Section 7. The absolute numbers are based on a count of the words with <ε> (not as part of the diphthongs <ει> and <ευ>) or <αι> without the relevant regularization of <ε> or <αι> in the Trismegistos Words database (1 February 2020) and a count of the interchanges in the relevant positions in the word in the Trismegistos Text Irregularities database (1 February 2020). The p-value is calculated in R with a Fisher's Exact test between the two directions of interchange in each context. Conventionally, p-values lower than 0.05 are considered enough to reject the null hypothesis that these two variables are independent of each other.

praesidium (garrison). It is spelt as πρε- in 11 per cent of these tokens, a frequency effect of the high contingency spelling <ε> in Greek in this context, clearly higher than the frequency effect of the use of the high contingency spelling <ε> in general (1.03 per cent). The opposite change, writing <πραι> instead of <πρε>, which would result in an uncommon spelling pattern in Greek, occurs only three times in the whole corpus (0.08 per cent), which is lower than the general interchange of <αι> instead of <ε> (0.18 per cent). There are some linguistic contexts in which a particular interchange indeed occurs significantly more often than on average. The use of <ε> instead of <αι> occurs more than twice as frequently (> 2.06 per cent) after <ι>, <λ>, <ξ>, <π>, <σ>, <ψ> and before <κ>, <ν>, <ρ>, <τ>, <ω> and the occurrence of <αι> instead of <ε> is twice as frequent than average (> 0.36) in the context after <κ>, <σ>, <τ> and before <α>, <δ> and word-finally.[7] Considering the diversity in phonetic contexts, the preference for these contexts is not easily explained by a phonologically conditioned sound merger or co-articulation effects only and there may be other reasons for the high frequency of interchanges in these contexts, which will be explored in the following sections.

There are only two linguistic contexts that form a clear exception to this general preference for the high contingency spelling <ε>. After <κ> and word-finally, the grapheme <αι> is more frequently used than <ε> and interchange of <αι> instead of <ε> is also more frequent than the other way around. Although these distributions confirm that frequency effects may play an important part in the choice of graphemes, this does not explain why <αι> is used more frequently instead of <ε> in precisely these two contexts and not in others. These issues will be studied in more detail in Sections 5 and 6.

4.2. Frequency of interchanges of <ο, ω>

Table 4 shows the same frequencies of interchanges <ο, ω> for different linguistic contexts. The of <ο, ω> occur in almost all contexts, so the loss of length distinction and the merger of the two phonemes do not seem to be conditioned by a specific linguistic context only, as has been concluded before (see introduction). In most of the contexts, the difference between the two types of interchanges is significant ($p < 0.05$). In most of these cases (first rows in the table), <ο> is more common than <ω> and also the interchange of <ο> instead of <ω> is more frequent than the other way around. This means that the writer could have opted to write <ο> because this grapheme is most common in general or in this particular context. For example, the spelling of the prefix προ- is much more frequent (94 per cent) than the same string with the low contingency vowel πρω- (6 per cent) and thus we also find a more than average number of examples of <ο> instead of <ω> in this string (3 per cent) than <ω> instead of <ο> (0.3 per cent).

There are some linguistic contexts in which the interchanges occur more often than on average. For the use of <ο> instead of <ω>, the instances after <π> and before <π>, <κ>, <λ>, <μ> occur more than twice as often (> 2.08 per cent) as this interchange on average, while <ω> instead of <ο> is more common (> 1.0 per cent) after <ε>, <υ>, <δ>, <χ>, <σ>.[8]

[7] Each of those has a p-value lower than 0.05 in a Fisher's Exact test, when the use of the two graphemes in this specific context is compared with the use of the same graphemes in all contexts. Testing and comparing the significance of each of the contexts individually in this way may potentially lead to occasional false positives (multiple comparisons problem). A stricter significance threshold could be used for the individual comparisons in order to compensate for this potential inference problem, but since the p-values are already extremely low for the contexts mentioned here (for most $p < 2.2e\text{-}16$ and for all $p < 2.4e\text{-}08$) correction for this does not seem to be required.

[8] Each of those has a p-value lower than 0.05 in a Fisher's Exact test, when the use of the two graphemes in this specific context is compared with the use of the same graphemes in all contexts. Testing and comparing the significance of each of the contexts individually in this way may potentially lead to occasional false positives (multiple comparisons problem). A stricter significance threshold could be used for the individual comparisons in order to compensate for this potential inference problem, but since the p-values are already extremely low for the contexts mentioned here (for most $p < 2.2e\text{-}16$ and for all $p < 7.4e\text{-}15$) correction for this does not seem to be required.

Again, considering the wide range of phonetic contexts, the preference for these contexts is not easily explained by a phonologically conditioned sound merger or co-articulation effects only and there may be other reasons for the high frequency of interchange in these contexts.

In some cases, the difference between the two directions is significant (see Table 4), but frequency effects would not be able to explain the preference for the direction of interchange. This applies to interchanges after <ι>, <ε>, <θ>, and before β, where <ο> is regularly written instead of <ω>, while <ω> is in fact more frequent in these contexts. The opposite is true for interchanges after <μ>, <υ>, <γ>, in which contexts <ο> is most common, but the use of <ω> instead of <ο> is most frequent. After <δ>, <χ>, <σ>, word-finally and before <ν> and <φ> there is no (significant) frequency effect explaining the difference between the two types of interchanges. In most of these situations, the proportional frequency of occurrence of <ο> and <ω> is relatively similar and the percentage of interchanges are comparable as well. The choice for <ο> in these cases could still be a frequency effect generated by the higher overall frequency of <ο>, but the choice for <ω> instead of <ο> in these cases would require another explanation. These issues are addressed in the following sections.

Table 4. Frequency of interchanges of <ο, ω> in different linguistic contexts.

Expected	<ω>			<ο>			
Written	<ο>	% <ο>	<ω>	<ω>	% <ω>	<ο>	p-value
Total	8,236	1.01	807,994	6,632	0.49	1,343,653	< 2.2e-16
Word-initial	507	1.03	48,834	623	0.53	116,925	< 2.2e-16
after π	274	2.14	12,512	229	0.18	127,672	< 2.2e-16
after φ	138	1.56	8,727	183	0.56	32,288	< 2.2e-16
after ρ	686	1.03	66,203	491	0.29	167,609	< 2.2e-16
after λ	438	1.21	35,906	303	0.39	77,041	< 2.2e-16
after ν	479	0.99	47,707	368	0.21	172,863	< 2.2e-16
after τ	1,186	0.70	167,539	1,174	0.58	202,295	1.629e-06
after κ	351	0.96	36,147	506	0.76	66,245	0.0006473
after -ι	861	0.66	129,664	336	0.30	111,511	< 2.2e-16
after ε	1,271	1.58	79,163	139	1.11	12,362	4.18e-05
after θ	246	1.26	19,267	91	0.91	9,873	0.007674
after μ	231	0.55	41,480	455	0.91	49,306	2.307e-10
after -υ	38	1.03	3,664	180	1.76	10,048	0.001908
after γ	125	0.71	17,452	188	0.92	20,270	0.02637
after δ	404	1.39	28,,605	461	1.22	37,292	0.05323
after χ	325	1.13	28526	327	1.28	25,181	0.09747
after σ	231	1.32	17,282	320	1.24	25,460	0.4848
before π	204	4.54	42,91	230	0.78	29,168	< 2.2e-16
before τ	485	1.35	35,444	304	0.71	42,771	< 2.2e-16
before κ	140	2.10	6,520	312	0.84	36,980	< 2.2e-16
before σ	1,241	0.93	132,518	676	0.18	382,635	< 2.2e-16
before λ	186	3.82	4,682	314	0.25	124,851	< 2.2e-16
before μ	1,377	3.56	37,309	832	0.68	120,681	< 2.2e-16
before ρ	584	1.23	46,896	491	0.79	61,846	2.334e-13
before δ	56	1.54	3,577	200	0.67	29,655	2.809e-07
before γ	29	1.06	2,716	207	0.45	45,715	8.906e-05
before β	147	0.51	28,726	26	0.11	23,309	< 2.2e-16
Word-final	881	0.77	113,967	750	0.74	100,518	0.4856
before ν	1,986	0.60	330,226	1,662	0.59	281,070	0.6172
before φ	65	0.74	8,765	116	0.76	15,158	0.8771

Note: Linguistic contexts with less than 100 interchanges for both directions are left out. The absolute numbers are based on a count of the words with <ο> (not as part of the diphthongs <ου> and <οι>) or <ω> without the relevant regularization of <ο> or <ω> in the Trismegistos Words database (1 February 2020) and a count of the interchanges in the relevant positions in the word in the Trismegistos Text Irregularities database (1 February 2020). The p-value is calculated in R with a Fisher's Exact test between the two directions of interchange in each context. Conventionally, p-values lower than 0.05 are considered sufficient to reject the null hypothesis that these two variables are independent of each other.

5. Lexical analogy

The quantitative results in section 4 have shown that, although both interchanges of <ε, αι> and <ο, ω> occur in many different contexts, the general distribution of the graphemes across these linguistic contexts may have had an impact on spelling choices and the direction of interchange. Some effects were observed that could not be explained by the proportional frequency only:

1 interchanges in some linguistic contexts are significantly more or less frequent than the interchange on average;
2 low contingency spellings which are more frequent than the high contingency spellings in a specific context may or may not lead to higher percentage of interchanges; and
3 low contingency spellings that are not the most frequent in a particular context can sometimes still be used more frequently than on average or even become more frequently used than high contingency spellings in that context.

In the current and following Sections (5–7), examples from these three groups will be examined further to see whether lexical and morphology analogy or changes in spelling practices may be able to explain these exceptions.

Analogy was introduced in section 2.3 as a potential influence on the selection of graphemes during phoneme-to-grapheme conversion. It was reported there that the source for analogy is likely to be a high frequency item and analogy is more likely to have a stronger effect on the use of low contingency spellings. This may be able to explain, for example, why interchanges in some contexts are more frequent than expected or why low contingency spellings are preferred above high contingency spellings in certain contexts.

One of the best examples of the possible impact of lexical analogy on the spelling variation in this study is the frequent use of <αι> instead of <ε> after <κ>. The high contingency spelling <ε> turned out to be preferred in almost all linguistic contexts (see Section 4.1), but still <αι> was used instead of <ε> more commonly than the other way around after <κ>. This may be related to the presence of this spelling pattern in one of the most frequent Greek lexemes: καί *kaí* 'and'. This lexeme represents 86 per cent of all attestations of the string <και> in the corpus. The variant spelling as <κε> occurs in only 0.15 per cent of these tokens, which is one of the lowest rates of interchange of <ε> instead of <αι> and this is a clear indication that the standard spelling of this basic word is likely to have been present in the orthographic lexicon of most writers. Based on the superficial similarity of the pattern, writers may also have chosen to write <αι> after <κ> in lexemes that are not present in their orthographic lexicon and are conventionally written as <κε>, resulting in the higher frequency of <αι> instead of <ε> spellings after <κ>. This seems particularly common at the beginning of the word, such as in καιφαλέου *kaiphaléou* for κεφαλαίου *kephaléou* 'capital sum-GEN.SG' (P.Oxy. LXXII 4897, 10; Oxyrhynchus, 391 CE), καιλεῦσε *kaileûse* for κελεῦσαι *keleûsai* 'order-INF' (P.Euphrates 2, 17; Syria, 245–248 CE), καίρδος *kaírdos* for κέρδος *kérdos* 'profit-NOM.SG' (P.Naqlun I 13-14, 3; Naqlun, VI CE) and καιραμίων *kairamíōn* for κεραμίων *keramíōn* 'vessel-GEN.PL' (PSI XII 1249, 47–48; Oxyrhynchus, 265 CE). The link with καί is also confirmed in more extreme cases where both <κ> and <αι> are newly introduced, such as καιωμετρίας *kaiōmetrías* for γεωμετρίας *geōmetrías* 'land survey-GEN.SG' (BGU XII 2157, 9; Hermopolis, 485 CE)

In some other cases, the high frequency of a low contingency spelling does not show an impact on the number of interchanges in a specific context. For example, the grapheme <ω> is more frequently attested than <ο> in the position before <β>, but still <ο> seems more commonly written instead of <ω> (0.51 per cent) than the other way around (0.11 per cent). In this case, it seems that one type of lexeme has in fact had a great impact on the higher number of spellings with <ο>. Up to 98 per cent of the interchanges of <ο> instead of <ω> before <β>

concern compounds with ὀβολός *obolós* 'obol' as a second part, such as διώβολον *diṓbolon* 'two-obol coin', τριώβολον *triṓbolon* 'three-obol coin', τετρώβολον *tetrṓbolon* 'four-obol coin' or ἡμιωβέλιον *hēmiōbélion* 'half-obol coin', and these compounds are conventionally spelt with a lengthened <ω> in Greek. These lexemes are often abbreviated (especially διώβολον *diṓbolon* and ἡμιωβέλιον *hēmiōbélion*) in the papyri, but when the first part is written out in full the spelling with <ο> is very common (20 up to 60 per cent of the tokens of the lexemes mentioned above). These variant spellings can be explained not only by the high contingency spelling of <ο> in general, but also by analogy to the spelling of the parts of the compound, such as ὀβολός *obolós*. Analogy to the individual parts is shown even more clearly in occurrences with multiple spelling deviations, such as ἡμιοβόλιον *hēmiobólion* instead of ἡμιωβέλιον *hēmiōbélion* based on ὀβολός *obolós* (O.Wilck. II 93, 6; Syene, 114 CE) and δυόβολον *duóbolon* for διώβολον *diṓbolon* based on δυό *duó* 'two' (O.Wilck. II 371, 2; Thebes, 35 CE). If we remove these compounds from the category 'before <β>', the (more varied) other uses of the high contingency spelling <ω> instead of <ο> are more frequent than <ο> instead of <ω> in this context, as one would expect based on the frequency of the string.

Another example of a higher frequency of a high contingency spelling in a context that can be influenced by lexical analogy is that interchanges of <ο> instead of <ω> are more frequent than expected in the contexts after <π> and before <λ>. The string <πολ> *pol* concerns 47 per cent of the interchanges after <π> and even 69 per cent of the interchanges before <λ>. The proportional distribution of the spelling <πολ> *pol* is much higher (95 per cent) with only 5 per cent of the tokens conventionally written as <πωλ> *pōl*. Almost half (44 per cent) of the attestations of <πολ> *pol* in the corpus concern the lexeme πόλις *pólis* 'city', and derived forms such as μητρόπολις *mētrópolis* 'capital city' and πολίτης *polítēs* 'citizen'. Interchanges of <ω> instead of <ο> occur only in 0.11 per cent of the tokens of these lexemes, while <ο> is written instead of <ω> in 6.2 per cent of the tokens of <πωλ> *pōl*, mostly involving the lexemes πωλέω *pōléō* 'to sell', πώλης *pṓlēs* 'seller' and derived forms. It seems therefore that the frequent and well-known lexeme πόλις *pólis* may have provided a basis for analogy for the spelling of the less frequently occurring lexemes πωλέω *pōléō* and πώλης *pṓlēs*. Compare, for example, the relatively frequent writing of <πολιν> *polin* 'for πωλεῖν *pōleîn* 'to sell-INF' (P.Mich. V 245; Tebtynis, 47 CE), where together with the variant spelling of the infinitive ending the interchange produces an identical spelling to the accusative singular of πόλις *pólis*.[9] We may even have some examples of lexical priming for this particular interchange. In three out of the five cases where a form of πόλις *pólis* 'city' is followed directly by a form of πώλης *pṓlēs* 'seller', both pronounced as ['pɔlis], the form of πώλης *pṓlēs* is spelt with <πολ>.[10] This is more frequent than the already high rate of misspellings of πώλης *pṓlēs* generally (11 per cent) which suggests that the cognitive effect of the spelling of πόλις *pólis* the word before could have had a priming effect on the writing of the lexeme πώλης *pṓlēs* following directly after.

Lexical analogy could also provide an explanation for the use of low contingency spellings that are not the most frequently used variant in the linguistic context and are therefore not easily explained in other ways. Just as we saw in the case of spelling πόλις *pólis* 'city' and

[9] Accents are not generally written in documentary papyri, which allows for an identical spelling of these two words, even though the stress accent is supposed to be on different syllables in the spoken language. The two forms should, therefore, not technically be regarded as homonyms, but it remains remarkable that the spelling with <ο> and <ι> is found in nine out of the fifteen examples of the interchange in the infinitive πωλεῖν.

[10] In P.Oxy. XIX 2230, 4 (119–124 CE; BL VIII, 255) ἱματιοπολ(ῶν) *himatiopol(ōn)* is written instead of ἱματιοπωλ(ῶν) *himatiopōl(ōn)* 'cloth-seller-GEN.PL' directly following πόλ(εως) *pól(eōs)* 'city-GEN.SG'; in P.Oxy. LXXII 4897, 5 (391 CE), ἀρτωπόλεις *artōpóleis* is written instead of ἀρτοπώλης *artopṓlēs* 'bread-seller-NOM.SG' directly following πόλει *pólei* 'city-DAT.SG' (same spelling used in l. 20 without directly priming); in P.Oxy. VIII 1133, 7 (396 CE) ὀπωροπόλῃ *opōropólēi* is written instead of ὀπωροπώλῃ *opōropṓlēi* 'fruit-seller-DAT.SG' directly following π]όλει *p]ólei* 'city-'. In P.Oxy. XXXI 2567, 6–7 (253 CE) φαρμα|κοπώλου 'drug-seller-GEN.SG' is written directly following πόλεως *póleōs* 'city-GEN.SG' and in P.Oxy. XLIII 3141, 9 and 17 (300 CE) ἐριοπώλης *eriopṓlēs* 'wool-seller-NOM.SG' is written also directly following πόλεως *póleōs* 'city-GEN.SG' in l. 9.

πώλης *pṓlēs* 'seller', homonyms are a likely subject of lexical analogy. The grapheme <ε> is the high contingency spelling and also frequently used instead of <αι> after <π> and before <δ>, but also <αι> is used instead of <ε> more frequently than average before <δ>. Almost all of the interchanges of <ε> and <αι> in the position between <π> and <δ> concern variant spellings of the lexemes παῖς *paîs* 'child', the diminutive παιδίον *paidíon* and other derivatives, and the lexemes πεδίας *pedías* 'field', πεδίον *pedíon* with the same meaning, and derivatives of those. The oblique forms of the lexeme παῖς *paîs* are attested 726 times in the corpus and are spelt with <ε> fifteen times (2 per cent), whereas the lexeme παιδίον *paidíon* occurs 828 times and is spelt with <ε> forty five times (5 per cent). The difference between the lexemes πεδίας *pedías* and πεδίον *pedíon* is even more extreme: πεδίας *pedías* occurs 371 times in the corpus, but is only once spelled with <αι> (0,3 per cent), while the lexeme πεδίον *pedíon* 467 in the corpus is spelled seventeen times with <αι> (4 per cent).[11] Especially in the case of the use of the low frequency spelling <αι>, the frequently occurring homonym παιδίον *paidíon*, which was probably (more) familiar to most writers, seems to have influenced the choice for <αι> in the less frequently occurring lexeme πεδίον *pedíon*. The influence of πεδίον *pedíon* on the higher number of instances of <ε> in παιδίον *paidíon* than in παῖς *paîs* is more difficult to ascertain, since it seems less likely that (many) writers would have been aware of the spelling of πεδίον *pedíon* and not of that of παιδίον *paidíon*, which is twice as frequently attested in the corpus. There could also have been other mechanisms influencing the choice of a high contingency spelling in those cases.

6. Morphological analogy

In the same way as other lexemes could have an influence on spelling choices, similarities between morphemes could also result in the variation between graphemes in morpheme position. For example, the generally low contingency spellings <αι> and <ω> are both more frequent than <ε> and <ο> in word-final position. The use of <ω> instead of <ο> is as frequent as <ο> instead of <ω> word-finally and the interchange of <αι> instead of <ε> is even more frequently attested word-finally than the other way around. Interchange in positions after many other letters, such as after <σ>, <τ>, <ε>, <ι>, <θ> and before <ν>, are also more frequent than average, because they typically involve morphemes. In this section I will compare the frequency and interchanges of graphemes in similarly sounding morphemes. Most of the case morphemes show clear frequency effects, see Table 5.

Table 5. Absolute number of morphemes (N) compared to the percentage of interchanges (%) of <ε, αι> and <ο, ω> in nominal endings.

Morpheme	N	% interchange	Morpheme	N	% interchange
NOM.PL -αι	82,882	0.3%	VOC.SG -ε	2,599	2.3%
NOM.PL -ες	14,342	0.2%	DAT.PL -αις	7,154	1.4%
DAT.SG -ῳ	77,752	0.5%	NOM-ACC.SG -ο	34,207	1.2%
GEN.PL -ων	218,320	0.5%	ACC.SG -ον	101,884	0.6%

The nominative plural ending -αι *-ai* of the first declension (N = 82,882) is more frequently attested than the vocative singular ending -ε *-e* of the second declension (N = 2,599) in our corpus and also the spelling of <αι> instead of <ε> for the vocative (2.3 per cent) is relatively more frequent than writing <ε> instead of <αι> for the nominative plural (0.3 per cent). The same applies to nominative plural ending of the third declension in -ες *-es*, which occurs more frequently than the dative plural of the first declension in -αις *-ais*, and -ες *-es* is also written more

[11] Both differences between the two lexemes are statistically significant with $p < 0.05$ in a Fisher's Exact test.

frequently for the dative plural (1.4 per cent) than -αις *-ais* for the nominative plural (0.2 per cent). Similarly in the second declension, the dative singular ending -ῳ *-ōi* is more frequent than the nominative-accusative singular neuter ending -ο *-o* and also <ω> is more frequently written for the neuter ending (1.2 per cent) than <ο> is used to write the dative singular (0.5 per cent). The effect is less visible for the most frequent interchange between the accusative singular -ον *-on* and genitive plural -ων *-ōn*, but also there the higher general frequency of the genitive plural ending in the corpus may be related to the more frequent spelling of -ων *-ōn* for the accusative singular (0.6 per cent) than -ον *-on* for the genitive plural (0.5 per cent).[12] Since the most frequent morphological interchanges mentioned above seem to involve changes across multiple dimensions at once, such as case, number, gender and declension, the distributional frequency of occurrence of the graphemes in a certain position (e.g. at the end of the word) is probably more important for these interchanges than their morphological or semantic properties. The same also applies to the very common spelling of the accusative singular personal pronouns as σαι *sai* (3.8 per cent) and (ἐ)μαι *(e)mai* (3.6 per cent) instead of σε *se* and (ἐ)μέ *(e)mé*, even though the ending -αι *-ai* cannot be used for an accusative nor a singular form in Greek and does not occur in the standard paradigm of the first and second person pronouns at all.

The relatively frequency of interchanges of in verbal endings is generally modest (0–2.5 per cent), but there are some endings for which interchange is (much) more common, see Table 6.

Table 6. Absolute number of morphemes (N) compared to the percentage of interchanges (%) of <ε, αι> in verbal endings.

Morpheme	N	% interchange	Morpheme	N interchange	%
3SG.M -ται	22,179	1.5%	2PL.A -τε	2,455	8.9%
INF.M -σθαι	12,970	1.2%	2PL.M -σθε	605	14.9%
PRES.IMP.A -ε	1,244	2.3%	2SG.AOR.IMP.M -(σ)αι	957	13.2%
3SG.A -ε	3,554	2.0%	AOR.INF.A -(σ)αι	17,950	2.1%

The spelling <αι> is frequently used in attestations of the second person plural active ending -τε (8.9 per cent) and middle ending -σθε *-sthe* (14.9 per cent). These interchanges could be explained by analogy to the much more frequently attested and similarly sounding endings of the third person singular middle –ται and the infinitive middle -σθαι *-sthai*. Again, the superficial similarity and the distributional frequency of the specific linguistic context could be considered the main motivation for this analogy. For some endings, multiple parallels can be drawn. The second person singular imperative aorist middle ending in -(σ)αι *-(s)ai* is not frequently attested and often spelt as -(σ)ε *-(s)e* (13.2 per cent) (see also Leiwo 2017, Dahlgren & Leiwo 2020). Morphological analogy to the second person singular imperative present active ending -ε *-e* seems most likely (Leiwo 2017), although influence from the final <ε> in the very frequent third person singular active ending cannot be excluded. The frequency of the aorist infinitive active ending -(σ)αι *-(s)ai*, on the other hand, seems to prevent very frequent interchange of this particular ending, keeping the percentage of interchanges much lower and very similar to those of the third person singular active ending -ε *-e* (2.1 and 2 per cent resp.), compare Example (2).

(2) δέξε [παρὰ] Νιλᾶτος πέντε εἰστατήρων
dékse [pará] Nilâtos pénte eistatḗrōn
reg. *déksai statḗrōn*
take-2.SG.IMP.AOR.M from Nilas-GEN.SG five stater-GEN.PL
'take five *staters*' worth from Nilas' (O.Did. 373, 3-4; ca. 88-96 CE)

[12] All these differences are statistically significant with a p-value < 0,05 in Fisher's Exact test.

In Example (2), the writer writes in direct speech with the aorist imperative δέξε *dékse* 'take' spelled with -ε, but uses the spelling -αι three times in the two aorist infinitives ἀγοράσαι *agorásai* 'to buy-AOR.INF.A' and πωλῆσαι *pōlēsai* 'to sell-AOR.INF.A' in lines 5 and 6 and in the more formulaic aorist imperative ἄσπασαι *áspasai* 'greet' in line 7. In direct speech, the writer may have been more inclined to stress the imperative value in more typical ways (in analogy to the imperative present active ending), while the aorist infinitive ending –σαι *–sai* looks familiar enough.

The spelling of <ο> instead of <ω> concerns, among others, several subjunctive endings, such as the active first person plural -ωμεν *-ōmen*, and the middle first person singular -ωμαι *-ō mai* and third person plural -ωνται *-ōntai*. Of course, the percentage of editorial regularisations of subjunctive instead of indicative in the present tense (2.7–4.7 per cent) partly depends on the interpretation of the editors and it would be best to analyse these instances in their linguistic context and compare them with the use of the other subjunctive endings in the same text. On the other hand, the subjunctive is not very frequently used in the papyri and the spelling of <ο> instead of <ω> seems especially common in these endings of the aorist subjunctive, see Table 7.

Table 7. Absolute number of morphemes (N) compared to the percentage of interchanges (%) of <ο, ω> in verbal endings.

Morpheme	N	%	Morpheme	N	%
1.PL.FUT.IND.A -ομεν	729	14.4%	1.PL.AOR.SUBJ.A -ωμεν	639	8.9%
1.SG.FUT.IND.M -ομαι	810	3.1%	1.SG.AOR.SUBJ.M -ωμαι	208	8.8%
3.PL.FUT.IND.M -ονται	305	0.7%	3.PL.AOR.SUBJ.M -ονται	198	6.6%

The difference between aorist indicative and subjunctive forms does not rely on the difference between <ο, ω> only, but also on the use of the augment in the indicative forms. Furthermore, the future tense should not have subjunctive endings at all. The interchange of indicative and subjunctive endings in the aorist and future tense could, therefore, be explained by the spelling of the morphemes rather than their syntactic usage in context only. The interchange in the aorist subjunctive is not entirely unexpected, because the spelling with <ο> is the high contingency spelling and the endings -ομεν, -ομαι, and -ονται are much more frequently attested, not only in the future forms mentioned in Table 7, but also in the forms of the much more familiar present indicative. Both frequency and analogy to the present indicative forms would explain these variant spellings.

More surprising perhaps, is the opposite effect: the frequent use of <ω> instead of <ο>, especially in the first person plural active ending -ομεν, both in future (14.4 per cent) and present (8.1 per cent) tense. That this issue was bothering scribes, can be noticed from the correction of εὔχομαι *eúkhomai* 'I pray' into εὔχωμαι *eúkhōmai* in P.Oslo inv. 361, 25 (first half II CE) by the later addition of the second loop of the *omega* on top of the beginning of the *mu*. At first sight, one would perhaps think that this interchange is also related to the confusion between indicative and subjunctive endings, but this is not very likely, since the subjunctive is not frequently used in the papyri and therefore does not provide a likely source for analogy.[13] Furthermore, while the spelling of <ο> instead of <ω> in the aorist subjunctive endings is also regularly attested in the middle endings (see Table 7), the use of <ω> instead of <ο> seems particularly frequent in the first person plural active ending -ομεν *-omen*, followed by the first person singular middle ending -ομαι *-omai*. I think that the very frequently used

[13] The proportional distribution between the first person plural active in aorist indicative and subjunctive is resp. 78 per cent and 22 per cent and present indicative and subjunctive even 92 per cent vs. 8 per cent.

and well-known first person singular ending -ω would provide a much more likely basis for analogy for these first person endings in both present and future tenses. A possible illustration of this type of reasoning could perhaps be found in the abbreviations ἔχω(μεν) *ékhō(men)* 'we have' and ἀπέχω(μεν) *apékhō(men)* 'we receive', used in receipts (e.g. P.Bagnall 63, 4; 28 CE, and P.Sijp. 38m, 3; 64 CE), in which the first person plural ending is written with the same vowel as the first person singular ἔχω *ékhō* or ἀπέχω *apékhō*, but with <ω> written above the line indicating that this is an abbreviated form. The same spelling is also found in unabbreviated spellings of the same verbs in similar receipts, for example in O.Berl. 33 (116 CE), where the raised *omega* is followed by the first plural ending written in full: ἀπέχ^{ω}μεν *apékh^{o}men*. In this way, the variant spelling could be understood as a form of analogical levelling in the active endings of the paradigm. The first person middle ending could also be analogical to the first person active ending, but he plural middle endings would be less likely to receive analogical levelling of this type, as the /o/ in the first singular, first and third plural middle endings are all traditionally spelled with the same grapheme <ο>.

Examples of analogical levelling may also be found in the nominal paradigms. Several nouns in the third declension have a nominative singular ending in -ωρ *-ōr* or -ων *-ōn*, e.g. αὐτοκράτωρ *autokrátōr* 'emperor' and γείτων *geítōn* 'neighbour', while the oblique cases in these paradigms are traditionally spelt with <ο>, such as αὐτοκράτορος *autokrátoros* 'emperor-GEN.SG' and γείτονος *geítonos* 'neighbour-GEN.SG'. The same applies to the active participle, which has a nominative singular masculine ending in -ων *-ōn* in the present and future, e.g. the nominative ὤν *ṓn* and genitive ὄντος *óntos* 'being', and an active participle in -ως *-ōs* in the perfect, e.g. the nominative εἰδώς *eidōs* and genitive εἰδότος *eidótos* 'knowing'. Even though the <ο> spelling would be the high contingency spelling in these contexts, the spelling <ω> is also produced in the oblique cases of these nouns and participles in higher frequencies than interchanges before <ντ>, <ρ>, <τ> in general.[14] The nominative singular seems a likely source of analogy for the spelling of the oblique cases as well as the regular declension pattern -ων, -ωνος *-ōn*, *-ōnos* used in several other nouns, mostly personal names. For example, in a list of names in P.Cair.Isid. 10 (311 CE), the genitive Εὐδαίμονος *Eudaímonos* is consistently spelt Εὐδέμωνος *Eudémōnos* with the high contingency spelling <ε> and with <ω> in the genitive ending of the patronymic, replicating the spelling pattern of other personal names (also present in the list), such as Ἥρων *Hḗrōn* with genitive Ἥρωνος *Hḗrōnos* (l. 121), Σαραπίων *Sarapíōn* with genitive Σαραπίωνος *Sarapíōnos* (ll. 195, 205) and Ἰσίων *Isíōn* with genitive Ἰσίωνος *Isíōnos* (ll. 141, 173), in which the vowel of the nominative is also used in the oblique cases in standard spelling.

Outside of personal names, the <ω> spelling in the oblique cases is less common in Greek standard spelling, but the <ω> in the nominative singular could still have served as a basis for analogy for writers, such as γεγραφώτος gegraphōtos instead of γεγραφότος 'written-PERF.PTC.GEN.SG' and ἡγεμώνος *hēgemṓnos* instead of ἡγεμόνος *hēgemónos* 'leader-GEN.SG' (P.Fam.Tebt. 15, 137–138; 114–115 CE) and παντοκράτωρα *pantokrátōra* for παντοκράτορα *pantokrátora* 'almighty-ACC.SG' and ὑπαρχώντων *huparkhōntōn* for ὑπαρχόντων *huparkhóntōn* 'belong to-PR.PTC.GEN.PL' (P.Mich. XIII 662, 12 and 59; 645 CE, BL XIII, 139). This type of analogical morphological interchange commonly occurs in in the documents of the *grapheion* of Tebtynis, see e.g. αὐτοκράτωρος *autokrátōros* for αὐτοκράτορος *autokrátoros* 'emperor-GEN.SG' (7 per cent of the tokens), γίτωνος *gítōnos* 'neighbour-GEN.SG' for γείτονες geítones 'neighbour-NOM.PL' (17 per cent), διελῆλυθώτος *dielēluthṓtos* for διελῆλυθότος *dielēluthótos* 'passed-PERF.PTC.GEN.SG' (40 per cent) and συνκυρώντων *sunkurṓntōn* for συγκυρόντων *sugkuróntōn*

[14] The differences between the number of interchanges in the endings of these nouns and participles and the interchanges before <ντ>, <ρ> and <τ> in general, resp. 0.9 vs. 0.3 per cent, 1.2 vs. 0.8 per cent and 1.3 vs. 0.7 per cent, are statistically significant with a p-value < 0.05. The interchanges before <ν> should not be compared as they also include the more frequent interchange between the accusative singular and genitive plural (see above).

'come together with-PR.PTC.GEN.PL' (65 per cent), and the practice may have become a matter of convention for some of these writers.

Modern research shows that children learning how to spell use their knowledge of the spelling of base forms in order to spell inflected forms and that this is related to their level of morphological awareness (Kemp 2006). That the spelling of the nominative case could have served as a basis for analogy can also be observed by comparing the number of interchanges for the nominative and oblique cases of a particular lexeme. For example, the common noun ἡγεμών *hēgemṓn* 'leader, prefect' and the personal name Εὐδαίμων *Eudaímōn* are always spelt with <ω> in the nominative singular (resp. 142 and 328 times) without any interchanges, while the oblique cases are spelt with <ω> instead of <ο> resp. 26 and 19 times in the corpus, which is 2.7 and 2.6 per cent of the attestations of oblique cases. Even though the nominative singular does not have a higher token frequency than the oblique cases, the use of the low contingency spelling <ω> in other cases could have been based on analogy with the <ω> nominative singular which seems to have been well known to most writers judging from the absence of interchanges in the nominative form. Just like the first person singular of verbal endings, the predominance of the nominative singular spelling can also be observed in the spelling chosen for abbreviated forms. For example, the spelling with <ω> is used in the oblique endings of τέκτων *téktōn* 'carpenter' (5.3 per cent) and κάνων *kánōn* 'standard' (5.6 per cent) both when they are written in full as well as in abbreviations, regardless of the function of the noun. Compare, for example, the abbreviation τέκτ^{ω} in BGU III 894 (109 CE), used both for the singular τέκτω(ν) α *téktō(n) 1* 'one carpenter' (nine times) and plural τέκτω(νες) β *téktō(nes) 2* 'two carpenters' (fifteen times). In CPR XIX 25 (VIII CE), the <ω> is also consistently written in oblique forms, both in more, e.g. κανώ(νος) *kanṓ(nos)* in l. 5, and less, e.g. κανώνο(ς) *kanṓno(s)* in l. 6, abbreviated versions of the noun.

7. Scribal conventions

In previous sections, I have shown how variation could emerge following the sublexical route of phoneme-to-grapheme conversion, influenced by other types of lexical and non-lexical knowledge and morphological awareness. In the case of multiple graphemes mapping onto one phoneme, the selection of a grapheme could be based on frequency of the grapheme (in general or in a specific context) or by analogy to more frequent and better known lexemes and morphemes with a similar pronunciation. Some lexemes show more variation in spelling than others. This may be because their exact spelling cannot easily be retrieved from the orthographic lexicon of most writers and because they contain low contingency spellings which are less straightforward to produce by phoneme-to-grapheme conversion. The same variant spelling of a lexeme could be generated by different writers individually following the sublexical route using the same cognitive principles. Although this may well have been the case for many of the examples of variant spellings discussed above, there are clear signs that variant spellings are not always produced by an *ad hoc* process of phoneme-to-grapheme conversion.

A good example of a variant spelling attested multiple times in a specific social context is the variant spelling of the lexeme μέτωπον *métōpon* 'forehead'. After the second reform under emperor Tiberian (14 CE), physical characteristics of all parties need to be noted down in contracts (Yiftach-Firanko 2008). This introduced the frequent use of particular words describing physical characteristics, such as μέτωπον *métōpon*, in the contracts drawn up in village writing offices, such as the one in Tebtynis.[15] The traditional spelling with <ω> only features in one of the documents preserved from the *grapheion* of Tebtynis until 26 CE (P.Mich. V 337, 26 CE), while the alternative (high contingency) spelling with <ο> is attested

[15] For more information about this archive see Trismegistos Archive ID 93.

in multiple documents, sometimes multiple times. After Kronion takes over the function of the head of the *grapheion* from his father Apion in 26 CE, variation continues and both forms come to be used. Unfortunately, descriptions of the physical characteristics are added by various scribes and they are usually not identified in the documents. This means that it is difficult to tell whether some scribes used one spelling consistently and others scribes another or whether they varied between them or copied from each other. At least one of the regular scribes of the *grapheion* seems to have memorised the spelling with <ο> and applies it consistently in the abstracts of registered contracts (P.Mich. II 121r, 42 CE). In this long document in four columns, the lexeme is fifteen times written in the same way: μέτοπ(-) *métop(-)* with the last letter <π> as an abbreviation marker on top of a small rounded *omicron*.[16]

Another lexeme, ἔλαιον *élaion* 'olive oil', also appears in an alternative spelling in many documents of the *grapheion*. Only three documents contain the traditional spelling with <αι> once (P.Mich. V 321, 42 CE, and duplicate contracts P.Mich. V 355 and PSI VIII 902, ca. 48-56 CE), whereas eight documents contain multiple occurrences of the (high contingency) spelling with <ε>. Most of these are accounts (P.Mich. II 123v, 127, 128) or registers of drawn up contracts (P.Mich. II 123r, 124, 128, V 238), possibly written in the hand of Kronion, the head of the *grapheion* at that time.[17] In these texts, Kronion (or a scribe always writing for him) consistently writes a second <ε> instead of <αι> in ἐλέου *eléou* 'olive oil-GEN.SG' and other derived words, such as ἐλεῶν(ος) *eleṓn(os)* 'olive yard-GEN.SG', ἐλεουργός *eleourgós* 'oil manufacturer' and ἐλεούργιον *eleoúrgion* 'oil press'. The consistent spelling by one writer in multiple documents suggests that the alternative spelling probably did not result from *ad hoc* conversion, but was present in the orthographic lexicon of the writer.

An alternative way of spelling a lexeme does not need to be limited to the orthographic lexicon of an individual writer, but could easily spread in the social context of a community of practice. The spread of an established variant spelling can be observed for the lexeme γεωμετρία *geōmetría* 'land tax'. There may be some isolated occurrences of the spelling γεομετρία *geometría* in the centuries BCE, but the spelling with <ο> only starts to gain popularity from the first century CE.[18] Variation is attested in tax receipts from southern Egypt, but in most places in the north the traditional spelling with <ω> is used.[19] An exceptional practice seems to have been established in Karanis in the environment of the tax collector Sokrates. The young Sokrates is first attested as a tax collector in 109 CE (Hagedorn 2008; Van Minnen 1994: 242).[20] In this receipt for taxes collected by him, the

[16] Two times an even shorter form is used: με(-) and μετ(-), resp. in col. III, iii and iv. In col. IV, i, l. 1 μετό(πῳ) is printed in the first edition (and online edition), but although the shape of the *omicron* and *pi* are not written very precisely, the last sign is clearly the *pi* of the abbreviation written above the line, similarly to the others.

[17] Apart from the accounts and registers possibly written by Kronion, also the abstracts of contracts in P.Mich. II 121r, col. IV, ii and a contract P.Mich. V 322a, 9 and 32 contain the spelling, resp. ἐλεῶν(ος), ἐλεούργιν and ἐλέου.

[18] The spelling may also be attested in BGU VI 1462, 2 and P.Heid. VI 382, 8 (dated to the third or second century BCE), but the readings are not certain. The first more secure reading is found in SB XXII 10942, 3 (4 BCE), a contract from the Oxyrhynchite nome. The first attestations in the Fayum seem to be in subscriptions to contracts found in the archive of the *grapheion* in Tebtynis (PSI VIII 918, 38–39 CE; P.Mich. V 272, 45–46 CE; see TM Archive ID 93) and in tax receipts from Elephantine (e.g. O.Wilck. 13–15, 54–59 CE) and Western Thebes (e.g. O.Heid. 154, 58 CE; O.Theb. 71, 67 CE).

[19] Attestations of the spelling <ο> are found in the south of Egypt around Thebes, Elephantine and Syene between the first and third century CE, but the spelling with <ω> also remains regularly attested in tax receipts from these areas. Individual attestations of the spelling γεομ(μετρία) also seem to be attested in other villages in the Fayum, such as Tebtynis (P.Mil.Vogl. VII 301, 143–144 CE), Philadelpheia (BGU VII 1607, II CE) and Theadelpheia (P.Horak 23, 148–149 CE; P.Berl.leigh. II 38, 150 CE), but in the contemporary tax register of Theadelpheia (P.Col. V.1-v-1b, 160 CE) the <ω> is part of the abbreviation, just like the earlier registers from the second century BCE in Kerkeosiris (e.g. P.Tebt. I 94, 2.14, 112 BCE).

[20] For more information on the archive see TM Archive ID 63 and Van Minnen (1994: 245).

spelling γεωμ(ετρίας) is still used, although an abbreviation ending in <ο> for γεο(μετρίας) seems to have been written in the previous receipt on the same roll (P.Mich. VI 383, 107–109 CE).[21] We do not know exactly how and where Socrates adopted the spelling with <ο>, but at the end of his career, more than sixty years later, he uses the spelling with <ο> consistently in his tax rolls (P.Mich. IV 223 + SB XIV 11710).[22] Together with the office of tax collector (Van Minnen 1994: 242), his way of spelling passed on to his son Sarapion, who writes γεομ(ετρίας) consistently in the rolls by his hand (e.g. P.Mich. IV 224 + 357a-c + 360, 172–173; 361b, 173–174).[23] The practice may have spread among some other colleagues as well, such as the tax collector Eudaimon in neighbouring village of Kerkesoucha, although it is difficult for us to follow the spread outside of known family archives.[24]

8. Idiolects: back to Gemellus

Let us now return to the spelling in the letters written by Bellenus Gemellus between 95 and 110 CE (P.Fay. 111–120), mentioned in Section 1. Bülow-Jacobsen (2001) suggested that the more common use of <αι> instead of <ε> and <ω> instead of <ο> in his letters was perhaps because Gemellus was more 'pretentious' than the writers of the ostraca or had a different background. Of course, Gemellus lived in a different area (the Fayum oasis, more towards the north of Egypt) and as a discharged army veteran and landowner he had a different social background than the writers in the Eastern desert. He was probably also older than the other writers (between the age of 63 and 78 when the letters were written), something which seems to have had an effect on his handwriting (Ast & Azzarello 2012). The letters are addressed to his estate manager Epagathus and his son Sabinus and this informal context probably did not require particular efforts to write more correctly than usual. All of this may help us to understand why we find a relatively high number of spelling variants, but they do not explain why precisely these variants or why he would have had a tendency to choose 'a more complex way of writing the sound' (Bülow-Jacobsen 2001: 160).

If we take the linguistic contexts of the variants into account, does Gemellus prefer low contingency spellings in his writing more than other writers? There are indeed no interchanges of <ε> instead of <αι> in these letters, but there are eleven examples of writing <αι> instead of <ε>. Out of these, seven examples (64 per cent) occur word-finally, in which <αι> is in fact the high contingency spelling and interchanges of <αι> instead of <ε> are also generally more common than the other way around (see Section 4.1). Three of these concern the accusative singular personal pronoun σαι instead of σε and the other four the verbal ending -ε, resp. of the third person singular ending (ἠγόρακαι *ēgórakai* for ἠγόρακε *ēgórake* 'to buy-3.SG.AOR' and ἤνενκαι *ēnenkai* for ἤνεγκε *ḗnegke* 'to bring-3.SG.AOR' in P.Fay. 119, 3 and 10) and of the second person singular imperative ending (γείνοσκαι *geínoskai* for γίνωσκε *gínōske*

[21] Other attestations of the variant in the Fayum also show a link to Karanis, e.g. receipt for taxes paid to Moros (SB XVI 12637, 125 CE) and Isidoros (BGU XIII 2287, 162 CE).

[22] The spelling with <ο> is used more than 600 times in these tax rolls from Karanis. The traditional spelling with <ω> is not used in these rolls, only reconstructed by editors in lacunae.

[23] The same spelling is also used in another tax roll from Karanis written in an unidentified hand (P.Mich. VI 362, 175 CE; see TM Archive ID 63), a tax register made up by an unknown tax collector in Karanis (SB XII 11115, 183-184 CE) and in various other tax receipts found in Karanis, such as those in the archive of the family of Gemellus Horion (P.Mich. VI 385, 182–185 CE; see TM Archive ID 90).

[24] See the spelling in the receipt for taxes collected by Eudaimon in Kerkesoucha found in the archive of the family of Gemellus Horion in Karanis (P.Mich. VI 384, 172 CE; see TM Archive ID 90) and the identical way of abbreviating the word in the roll with land tax receipts collected by Eudaimon and others in Kerkesoucha dated between 182–216 (P.Sijp. 39).

'to know-2.SG.IMP' in P.Fay. 117, 3).[25] These are all morphological contexts in which the use of <αι> instead of <ε> is more frequent than on average (see Section 6). It is perhaps not a coincidence that these verbs also occur in the same texts, which do not contain any other verbs ending in -ε, whereas he does use the spelling -ε for third person singular verb endings in other letters (P.Fay. 111, 113, 114).

The direct context also seems to be important for the choice of spelling of the accusative singular of the personal pronoun. Gemellus' choice for the spelling σαι *sai* could be understood as a priming effect by the ending in -αι of immediately preceding verbs in all three instances, see μένφομαί σαι (for μέμφομαί σε) μεγάλως *ménphomaí sai* (for *mémphomaí se*) *megálōs* 'I blame you greatly' (P.Fay. 111, 3, and P.Fay. 112, 14) and ἐρρῶσθαί σαι (for σε) εὔχομαι *errṓsthai sai* (for *se*) *eúkhomai* 'I pray that you are healthy' (P.Fay. 117, 27). When there is no <αι> immediately preceding, the accusative σε is spelt with <ε> in two other letters, namely P.Fay. 116 and P.Fay. 118.[26] The accusative first person singular pronoun με is always spelt with <ε>, though both times occurring immediately after a verb ending in -ε, see Ἑρμοναξ ἐρώτησέ με *Hermonaks erṓtēsé me* (P.Fay. 113, 6) and ἐρώτησέ με Ἑρμο|ναξ *erṓtēsé me Hermonaks* (P.Fay. 114, 8–9), both meaning 'Hermonax asked me'.[27]

The interchanges of <ο, ω> are attested in both directions in the letters written by Gemellus, but the positions seem to be largely mutually exclusive. The spelling of <ο> instead of <ω> is attested word-medially in 94 per cent of the cases, while the majority of the variant spellings <ω> instead of <ο> is found word-finally (54 per cent) and word-initially (22 per cent).[28] The general preference for <ω> word-finally is also sometimes extended to non-inflected words, such as διώ *diṓ* for the conjunction διό *dió* 'wherefore' (P.Fay. 112, 14) and δύω *dúō* for the numeral δύο *dúo* 'two' (P.Fay. 111, 4, 12–13). Since <ω> can be regarded the high contingency spelling in word-final position and <ο> in most word-medial positions (see 4.2), most of the variants used by Gemellus are expected based on frequency. He consistently writes <ω> for the nominative singular article ὁ and relative pronoun ὅ and τώ for the neuter accusative singular τό. Both <ω> and

[25] In P.Fay. 119, 8–10, the situation is more complicated due to a mixed use of cases: Σαβῖνον τὸν Ψελλὸς τὸν | ἀπὸ Ψινάχεος τὸν μετ' ἐσοῦ | εἰς πόλιν ἤνενκαι ἐπιστολὴν *Sabînon tón Psellós tón apó Psinákheos tón met' esoû eis pólin ēnenkai epistolēn* 'Sabinus-ACC.SG the-ACC-SG Psellos-NOM.SG the-ACC.SG from Psinachis-GEN.SG the-ACC.SG with you-GEN.SG to city-ACC.SG to bring-3SG.AOR letter-ACC.SG', which translates as follows 'Sabinus, son of Psellos, of Psinachis, who is with you, brought a letter to the city'. The subject (Sabinus) is written mostly in the accusative instead of the nominative case, which would make one wonder whether he was thinking of an accusative and infinitive construction with the <αι> ending of the aorist infinitive, although there is an augment and no finite verb on which this infinitive would depend. The sentence is in principle constructed in the same way as the previous sentence of which the subject is in the nominative. Perhaps a preceding γίνωσκε *gínōske* 'know that', as he also uses in P.Fay. 117, could have been on the mind of the scribe?

[26] See P.Fay 116, 19-21: αἰὰν | [ἀπο]χωρῶι, πέμσωι (for ἐὰν | [ἀπο]χωρῶ, πέμψω) πρὸς | [σὲ ἵ]να σε ἀσπάσωμαι *aián apokhōrṓi, pémsōi* (for *eán apokhōrṓ, pémpsō*) *prós sé hína se aspásōmai* 'if I leave, I will send to you to greet you' and P.Fay. 118, 25–26: ἀσπάζου | τοὺς φιλοῦντάς σε πάντες (for πάντας) *aspázou toús philoûntás se pántes* (for *pántas*) 'greet all who love you'.

[27] The remaining four instances concern the spelling αιάν instead of ἐάν, which Gemellus spells with <ε> in two other letters (P.Fay. 111, 16, and 118, 22). Although Gemellus is not the only one occasionally using this low contingency spelling (thirty five times attested in the corpus), the frequency of interchanges in this lexeme is lower than average (0.5 per cent) and not easily explained by frequency or analogy, but the internal variation in spelling shows that the spelling could have been elicited by something in the direct context or perhaps other written language exposed to him.

[28] A comparable pattern can be found in a letter from Probus (P.Oxy. XIV 1683, late IV CE). The majority of the spellings of <ω> instead of <ο> are attested in high frequency positions word-initially, word-finally and before <ν> (11 out of the 19). Two others are in the first person singular middle (εὔχωμαι *eúkhōmai* 'I pray', l. 4) and first person plural indicative ending of verbs (ἔχωμεν *ékhōmen* 'we have', l. 13). He spells ἀπό *apó* 'from' consistently ἀπώ *apṓ* word-finally (ll. 21, 31), but also as a preverb, such as in ἀπωλάβῃς *apōlábēis* 'may you receive' (l. 7) and ἀπώδως *apṓdōs* 'may you give back' (l. 15). After ἀπώ, ἀπώδως is also spelled with <ω>, although he spells it δός *dós* 'you give' when it occurs separately in the letter (ll. 17, 20, 28). The accusative singular of the third declension is spelled as -ων *-ōn* (same as the genitive plural ending), e.g. μαρτυρων *marturōn* for μάρτυρον *márturon* for μάρτυρα *mártura* (l. 14) and λεβιτων *lebitōn* for λέβιτον *lébiton* for λέβητα *lébēta* (l. 22), by analogical levelling to the accusative singular of the second declension in -ον *-on* and frequency spelling of the <ω> before <ν>.

<ο> are very frequent in the position before <ν> and interchanges are found in equal amounts generally (see Section 6). Gemellus writes almost consistently <ο> instead of <ω> before <ν>, but the general confusion between the two shows sometimes in the accompanying articles, see P.Fay. 111, 24–26.[29]

Interchanges in verbal endings are less frequent. He opts for <ο> instead of <ω> in verbal endings two times, see διὸ γράφο συ εἴν[α] (for γράφω σοι ἵν[α]) *dió grápho su heína* (for *gráphō soi hína*) 'therefore I write you to' (P.Fay. 117, 19), after διό ending in <ο>, and also at least once after ἵνα *hína*.[30] He also writes three times <ω> instead of <ο> in the first person plural and the first person singular middle ending (see also Section 6), each time in subordinate modal clauses dependent on ἐπεί 'since', see ἐπὶ (for ἐπεὶ) βουλεύωμαι (instead of βουλεύομαι) | [εἰς π]όλιν ἀπελθῖν (for ἀπελθεῖν) *epí* (for *epeí*) *bouleúōmai* (instead of *bouleúomai*) *eis pólin apelthîn* (for *apeltheîn*) 'since I am intending to go to the city' (P. Fay. 116, 9–10), and ἐπὶ | χρίαν (for ἐπεὶ χρείαν) αὐτοῦ ἔχωμον (instead of ἔχομεν) *epí khrían* (for *epeí khreían*) *autoû ékhōmon* (instead of *ékhomen*) 'since we want to make use of him' (P.Fay. 117, 7–8) and ἐπὶ (for ἐπεὶ) κ[ο]|λάζωμαι (instead of κ[ο]|λάζομαι) αὐτῶν εἰς Ἀφρο|δίτην πόλιν *epí* (for *epeí*) *kolázōmai* (instead of *kolázomai*) *autṓn eis Aphrodítēn pólin* 'since I am in need of them at Aphroditopolis' (P.Fay. 120, 5–7). Apart from following general frequency patterns, Gemellus may have developed a few internal patterns in his use of spelling.

Based on this short overview, I think we can conclude that the variant spellings used by Gemellus are less exceptional than they might seem at first. His preference for spellings such as <αι> and <ω> is not random, but attested in specific lexemes, morphemes and linguistic contexts, mostly in agreement with generally higher frequencies of interchange. Gemellus does not opt for typical high contingency spellings, such as <ο> and <ε>, all the time, as the writers of the ostraca in Krokodilo seem to prefer (and Greek children still do nowadays in early stages of spelling acquisition, see Chliounaki & Bryant 2003), but he tries to use his knowledge of the frequency and distribution of Greek phoneme-to-grapheme mappings to adapt his selection of graphemes. Several alternative standards are applied consistently, but his choices are also often influenced by the direct linguistic context. This practice does not seem to be out of the ordinary in the informal writing of an elderly man with probably many years of writing experience and regular exposure to written language, but a long time since any form of formal education or training. A similar pattern is found in modern studies comparing spelling of children with that of adults with basic literacy. The adults tend to have better skills recognising orthographic patterns, as Gemellus seems to have, but more difficulties with derivational and inflectional morphemes (Worthy & Viise 1996). Morphology is the spelling domain that takes most time to master in Modern Greek (Protopapas et al. 2013). Similar tendencies could perhaps be identified in the writing of other regular writers in Roman Egypt who may have been aware of the potential difficulties of Greek spelling, but without enough (recent) education to master them completely.

[29] Note the interchange between -ον *-on* (ACC.SG) and -ων *-ōn* (GEN.PL) in καὶ τῶν στί|χον τὸν φυτον τῶν | ἐν τῷ προφήτῃ πότισον *kaí tōn stíkhon tón phuton tōn en tōi prophētēi pótison* 'and the-GEN.PL row-ACC.SG the-ACC.SG tree-ACC.SG the-GEN.PL in the-DAT.SG prophet-DAT.SG water-2.SG.IMP.AOR', meaning 'and water the row of trees at "the prophet"'.

[30] See P.Fay. 119, 22–23: εἵνα αἰὰν μέλλω πρὸς αὐτὸν | λογεῖν ἐχο (instead of ἔχω) αὐτά *eína aián méllō prós autón logeîn ékho* (instead of *ékhō*) *autá* 'in order to have them (i.e. the keys), if I am about to settle accounts with him'. Another example in P.Fay. 116, 17–18 was read εἵνα (for ἵνα) τῶι | [ἀδελ]φῶι πέμσομον (instead of πέμψωμεν) *heína* (for *hína*) *tṓi adelphōi pémsomon* (instead of *pémpsōmen*) 'in order that we may send (them) to my brother' in the *edition princeps*, but it looks like the *omega* could very well be written in πέμσωμον *pémsōmon* (based on the image online).

9. Conclusion and discussion

In this article, I applied a dual-route model for spelling to interpret two particular spelling interchanges often regularised by the editors of documentary papyri. First of all, it can be observed that the general frequencies, even of relatively frequently observed interchanges of <ε, αι> and <ο, ω>, are low (ca. 1 per cent). Most writers thus seem to know and use standard spellings most of the time. This would not be expected when the writers had very little knowledge of Greek orthography or no knowledge at all. In that case, phoneme-to-grapheme conversion would be the main route for spelling and, given the sound mergers taking place in the Greek language at the time, a large amount of spelling variation is to be expected.

The attested variation in spelling, on the other hand, is not randomly spread and there are clear frequency effects and distributional differences in the choice of graphemes. The high contingency spellings, the phoneme-to-grapheme mapping that is most frequent for that phoneme, namely <ε> and <ο>, are clearly more frequently used than low contingency spellings <αι> and <ω>. Low contingency spellings can still be used more than average – or even more than the generally high contingency spellings – in specific linguistic contexts, such as word-initially, word-finally or before or after specific graphemes, in which contexts they may even technically be the high contingency spelling. This type of general knowledge of the Greek written language, such as grapheme frequency and distributional pattern probabilities, is applied by many writers to spell words that they cannot retrieve from their orthographic lexicon. Low contingency graphemes can also be preferred in specific contexts based on analogy with a (more) frequent source, such as the spelling of <αι> after <κ> based on the generally well-known spelling of the highly frequent lexeme καί *kaí* 'and'. In some cases, the source for analogy is even attested in the direct context of use, suggesting that some form of priming may have had an impact on the chosen spelling. Like lexical analogy, morphological analogy, or general knowledge about the distribution of graphemes in morphemes, can also be used, for example to produce morphologically plausible spellings. In many cases however, the correct spelling of Greek morphemes, just like in modern-day Greek, requires morphological awareness, insight into the paradigms, inflections and derivational patterns and their exceptions rather than just applying more superficial knowledge about the distribution of graphemes in word endings.

Applying this new approach to the Example (1) discussed in the introduction, we can now observe that the general high contingency spelling <ο> is used to produce the correct spelling ἐπομνύμενος *epomnúmenos* 'swearing', while the low contingency spelling <ω> is preferred in θεὸν τὸν παντωκράτωρα *theòn tòn pantōkrátōra* instead of the standard spelling θεὸν τὸν παντοκράτορα *theòn tòn pantokrátora* 'God the Almighty'. Before <ν> /*n*/, the proportional frequency of <ω> is almost equal to <ο>, and even slightly higher (see Section 4.2), as well as being morphologically attested in that position (-ων is a frequently occurring genitive plural ending). Therefore, confusion between -ων -ōn (genitive plural) and -ον -*on* (accusative singular) is very common in documentary papyri (see also Section 8). The use of <ω> continues in παντωκράτωρα *pantōkrátōra* for παντοκράτορα *pantokrátora* 'almighty'. The replacement of the first <ο> by <ω> would require comparison with the most commonly attested spellings in the declension of πᾶς, παντός *pâs, pantós* 'all' and its use as the first element in derivational compounds, but the use of the <ω> in the ending -ωρα -*ōra* can now be analysed as a case of morphological levelling based on analogy with the nominative singular ending in -ωρ -*ōr* (see Section 6). That the writer (or someone correcting him) was aware of his difficulties with this inflectional paradigm, can also be observed from line 4, where ἄρχωντος *árkhōntos* (produced in analogy with the nominative singular ἄρχων *árkhōn*) was later corrected to ἄρχοντος *árkhontos* 'prefect-GEN.SG'.

Variant spellings generated by sublexical phoneme-to-grapheme conversion are thus rarely random. The cognitive process of spelling is guided by patterns of frequency and probability

and can be influenced by lexical or morphological analogies. In which lexemes and morphological endings the variation occurs is hardly random either. Even though these spellings could all be generated *ad hoc* in the same way by various writers coincidentally, they could eventually also become part of the orthographic lexicon, or idiolect, of a specific writer or even spread from one scribe to another by shared scribal practices.

A lot of these variant spellings are attested in social contexts where one could also suspect a higher level of influence from Egyptian in the Greek spoken language. It is important to note, however, that these contexts do not only tend to show a higher level of particularly Egyptian-influenced spelling interchanges, but also a higher level of spelling variation in general. The spelling variants discussed in this article, which are also widely spread outside of Egypt, can hardly be the result of language contact only and explanations involving typical patterns in the Greek language are warranted. It seems more likely that these particular places accumulated a higher number of variant spellings because Greek standard spelling was less known or found to be less important to the local writers and alternative spellings based on basic knowledge of the Greek written language spread relatively easily in these environments.

Correspondence
joanne.stolk@ugent.be

References

For papyrus editions see Checklist of Editions of Greek, Latin, Demotic, and Coptic Papyri, Ostraca, and Tablets, at www.papyri.info/docs/checklist.

For TM numbers see www.trismegistos.org.

Apel, Kenn, Victoria S. Henbest & Julie Masterson, 2019. 'Orthographic knowledge: clarifications, challenges, and future directions', *Reading and Writing* 32(4), 873–889.

Arppe, Antti, Gaëtanelle Gilquin, Dylan Glynn, Martin Hilpert & Arne Zeschel, 2010. 'Cognitive corpus linguistics: Five points of debate on current theory and methodology', *Corpora* 5(1), 1–27.

Ast, Rodney & Giuseppina Azzarello, 2012. 'A Roman veteran and his skilful administrator: Gemellus and Epagathus in light of unpublished papyri'. in Paul Schubert (ed.), *Actes du 26e Congrès international de papyrologie, Genève, 16–21 août 2020*. Genève: Librairie Droz. 67–71.

Bar-On, Amalia & Victor Kuperman, 2019. Spelling errors respect morphology: A corpus study of Hebrew orthography, *Reading and Writing* 32(5), 1107–1128.

Barry, Christopher, 1994. 'Spelling routes (or roots or rutes)'. in Gordan D.A. Brown & Nick C. Ellis (eds.), *Handbook of spelling: Theory, process and intervention*. Chichester: Wiley. 27–49.

Barry, Christopher & Philip H. K. Seymour, 1988. 'Lexical priming and sound-to-spelling contingency effects in nonword spelling', *The Quarterly Journal of Experimental Psychology* 40A(1). 5–40.

Behrens, Heike, 2017. 'The role of analogy in language processing and acquisition'. in Marianne Hundt, Sandra Mollin & Simone E. Pfenninger (eds.), *The changing English language: Psycholinguistic perspectives*. Cambridge: Cambridge University Press. 215–239.

Bonin, Patrick, Alain Méot, Séverine Millotte & Christopher Barry, 2013. 'Individual differences in adult handwritten spelling-to-dictation', *Frontiers in Psychology* 4(202), 1–11.

Bosse, Marie-Line, Sylviane Valdois & Marie-Josèphe Tainturier, 2003. 'Analogy without priming in early spelling development', *Reading and Writing* 16(7). 693–716.

Bresnan, Joan, 2007. 'Is syntactic knowledge probabilistic? Experiments with the English dative alternation', in Sam Featherston & Wolfgang Sternefeld (eds.), *Roots: Linguistics in search of its evidential base*. Berlin: Mouton de Gruyter. 77–96.

Bucking, Scott, 2007. 'On the training of documentary scribes in Roman, Byzantine, and early Islamic Egypt: A contextualized assessment of the Greek evidence', *Zeitschrift für Papyrologie und Epigraphik* 159, 229–247.

Bülow-Jacobsen, Adam. 2001. 'The pronunciation of Greek in the Ostraca from the Eastern desert', in Isabella Andorlini, Guido Bastianini, Manfredo Manfredi, Giovanna Menci (eds), *Atti del XXII congresso internazionale di Papirologia, Firenze, 23–29 agosto 1998*. Firenze: Istituto Papirologico "G. Vitelli". 157–162.

Bybee, Joan, 2010. *Language, usage and cognition*. Cambridge: Cambridge University Press.

Chliounaki, Kalliopi & Peter Bryant, 2003. 'Choosing the right spelling in Greek: Morphology helps', *Revue Française de Linguistique Appliquée* 8(1). 35–45.

Clackson, James, 2015. 'Latinitas, Ἑλληνισμός and Standard Languages', *Studi e Saggi Linguistici* LIII (2). 309–330.

Dahlgren, Sonja & Martti Leiwo. 2020. 'Confusion of mood or phoneme? The impact of L1 phonology on verb semantics', in Ilja Serzant & Dariya Rafiyenko (eds), *Postclassical Greek: Contemporary approaches to philology and linguistics*. [Trends in Linguistics. Studies and Monographs 335]. Berlin: De Gruyter, 283–302.

Delattre, Marie, Patrick Bonin & Christopher Barry, 2006. 'Written spelling to dictation: Do irregularity effects persist on writing durations', *Journal of Experimental Psychology: Learning, Memory, and Cognition* 32(6), 1330–1340.

Depauw, Mark & Joanne Stolk, 2015. 'Linguistic variation in Greek papyri. Towards a new tool for quantitative study', *Greek Roman and Byzantine Studies* 55(1), 196–220.

Dixon, Maureen & Zofia Kaminska, 1997. 'Is it misspelled or is it mispelled? The influence of fresh orthographic information on spelling', *Reading and Writing* 9(5-6), 483–498.

Ellis, Nick C., 2017. 'Chunking in language usage, learning and change', in Marianne Hundt, Sandra Mollin & Simone E. Pfenninger (eds.), *The changing English language: Psycholinguistic perspectives*. Cambridge: Cambridge University Press. 113–147.

Evans, Trevor V. 2012. 'Linguistic and stylistic variation in the Zenon Archive', in Martti Leiwo, Hilla Halla-aho and Marja Vierros (eds), *Variation and change in Greek and Latin*. Helsinki [Papers and Monographs of the Finnish Institute at Athens 17]. 25–42.

Gentner, Dedre, 1983. 'Structure-mapping: A theoretical framework for analogy', *Cognitive Science* 7(2). 155–170.

Georgiou, George K., Minna Torppa, George Manolitsis, Heikki Lyytinen & Rauno Parrila, 2012. 'Longitudinal predictors for reading and spelling across language varying in orthographic consistency', *Reading and Writing* 25(2), 321–346.

Gignac, Francis Thomas. 1976. *A grammar of the Greek papyri of the Roman and Byzantine periods. Volume I: Phonology*. Milan: La Goliardica.

Grainger, Jonathan & Johannes C. Ziegler, 2011. 'A dual-route approach to orthographic processing', *Frontiers in Psychology* 2(54). 1–13.

Hagedorn, Dieter, 2008. 'Sokrates und Asklepiades, Praktoren in Karanis', *Zeitschrift für Papyrologie und Epigraphik* 167, 149–150.

Hasenäcker, Jana, Elisabeth Beyersmann & Sascha Schroeder, 2016. 'Masked morphological priming in German-speaking adults and children: Evidence from response time distributions', *Frontiers in Psychology* 7(929). 1–11.

Hilpert, Martin, 2017. 'Frequencies in diachronic corpora and knowledge of language', in Marianne Hundt, Sandra Mollin & Simone E. Pfenninger (eds.), *The changing English language: Psycholinguistic perspectives*. Cambridge: Cambridge University Press. 49–68.

Horrocks, Geoffrey, 2010. *Greek: A history of the language and its speakers*. 2nd edn. Chichester: Wiley-Blackwell.

Hundt, Marinne, Sandra Mollin & Simone E. Pfenninger (eds.), 2017. *The changing English language: Psycholinguistic perspectives*. Cambridge: Cambridge University Press.

Kemp, Nenagh, 2006. 'Children's spelling of base, inflected and derived words: Links with morphological awareness', *Reading and Writing* 19(7), 737–765.

Leiwo, Martti, 2017. 'Confusion of moods in Greek private letters of Roman Egypt?' in Mark Janse, Klaas Bentein & Jorie Soltic (eds.), *Linguistic variation and change: Tense, aspect and modality in Ancient Greek*. Leiden: Brill. 242–260.

Mair, Christian, 2017. 'From priming and processing to frequency effects and grammaticalization? Contracted semi-models in present-day English'. in Marianne Hundt, Sandra Mollin & Simone E. Pfenninger (eds.), *The changing English language: Psycholinguistic perspectives*. Cambridge: Cambridge University Press. 191–212.

Tainturier, Marie-Josèphe & Brenda Rapp, 2001. 'The spelling process', in Brenda Rapp (ed.), *The handbook of cognitive neuropsychology*. Philadelphia, PA: Psychology Press. 263–289.

Martin, Daisy H. & Christopher Barry, 2012. 'Writing nonsense: the interaction between lexical and sublexical knowledge in the priming of nonword spelling', *Psychonomic Bulletin & Review* 19(4), 691–698.

Mayser, Edwin & Hans Schmoll. 1970. *Grammatik der griechischen papyri aus der Ptolemäerzeit. Band I. Laut- und Wortlehre. I. Teil: Einleitung und Lautlehre*. Berlin: De Gruyter.

Perry, Conrad, Johannes C. Ziegler & Max Coltheart, 2002. 'How predictable is spelling? Developing testing metrics of phoneme-grapheme contingency', *The Quarterly Journal of Experimental Psychology* 55A(3). 897–915.

Protopapas, Athanassios, Aikaterini Fakou, Styliani Drakopoulou, Christos Skaloumbakas & Angeliki Mouzaki, 2013. 'What do spelling errors tell us? Classification and analysis of errors made by Greek schoolchildren with and without dyslexia', *Reading and Writing* 26(5). 615–646.

Protopapas, Athanassios & Eleni Vlahou, 2009. 'A comparative quantitative analysis of Greek orthographic transparency', *Behavior Research Methods* 41(4). 991–1008.

Stenroos, Merja, 2018. 'From scribal repertoire to text community. The challenge of Variable writing systems', in Jennifer Cromwell & Eitan Grossman (eds.), *Scribal pepertoires in egypt from the New Kingdom to the Early Islamic Period*. Oxford: Oxford University Press, 20–40.

Stolk, Joanne Vera, 2018. 'Encoding linguistic variation in Greek documentary papyri. The past, present and future of editorial regularization', in Nicola Reggiani (ed.), *Digital papyrology II: Case studies on the digital edition of Ancient Greek Papyri*. Berlin: De Gruyter, 119–137.

Stolk, Joanne Vera, 2020. 'Post-Classical Greek from a scribal perspective: Variation and change in contemporary orthographic norms in documentary papyri', *Mnemosyne* 73(5), 750–774.

Stolk, Joanne Vera, Ágnes Mihálykó, & Céline Grassien. in preparation. 'Egyptians hearing Greek: explaining non-standard orthography in liturgical texts from Thebes', in Anastasia Maravela and Ágnes Mihálykó (eds), *New Perspectives on Religion, Education, and Culture at Christian Western Thebes (VI–VIII)*.

TAINTURIER, MARIE-JOSÈPHE, MARIE-LINE BOSSE, DANIEL J. ROBERTS, SYLVIANE VALDOIS & BRENDA RAPP, 2013. 'Lexical neighbourhood effects in pseudoword spelling', *Frontiers in Psychology* 4(862). 1–8.

TEODORSSON, SVEN-TAGE, 1977. *The phonology of Ptolemaic Koine*. Gothenburg: Acta Universitatis.

VAN MINNEN, PETER, 1994. 'House-to-house enquiries: An interdisciplinary approach to Roman Karanis', *Zeitschrift für Papyrologie und Epigraphik* 100(••). 227–251.

WORP, KLAAS, 2013. 'Papyrology'. in Georgios K. Giannakis (ed.), *Encyclopedia of Ancient Greek language and linguistics*. Brill Online <https://doi.org/10.1163/2214-448X_eagll_COM_00000269>.

WORTHY, JO & NEVA M. VIISE, 1996. 'Morphological, phonological, and orthographic differences between the spelling of normally achieving children and basic literacy adults', *Reading and Writing* 8(2), 139–159.

YIFTACH-FIRANKO, URI, 2008. 'Who killed the double document in Ptolemaic Egypt?', *Archiv für Papyrusforschung* 54 (2), 203–218.

YIN, LI, R. MALATESHA JOSHI, DAOXIN LI & SEON-KEE KIM, 2020. 'Decisions about consonant doubling among non-native speakers of English: Graphotactic and phonological influences', *Reading and Writing* 33(7), 1839–1858.