Scribes and Language Use
in the Graeco-Roman World

edited by

Sonja Dahlgren, Martti Leiwo and Marja Vierros

Scribes and Language Use
in the Graeco-Roman World

Scribes and Language Use
in the Graeco-Roman World

edited by
Sonja Dahlgren, Martti Leiwo and Marja Vierros

Contents

Introduction

This volume examines the interface between individuals who plan and compose various documentary texts and the language used in those texts by analysing the complex processes of document creation and finalisation. It highlights the importance of variation across multiple linguistic dimensions—such as language choice, script, orthography, syntax, and document format—demonstrating how these factors interact to convey different social and functional meanings in non-literary writing.

The contributors focus particularly on scribes and their influence on the linguistic outcomes of documentary texts: What kind of language is written, and why? Who authored the text, and who physically wrote it? It is argued that research on the language of ancient and medieval non-literary texts must consider the scribal level alongside the edited text. By 'scribal level', we refer to (1) the design of the text itself and (2) its actual writing, i.e. the practical skills involved in a specific context. This approach also considers the choice of language and writing system, as the scribe's linguistic competence could significantly influence language selection.

In addition to detailed linguistic analyses, an important outcome of this volume is its exploration of the varied ways in which the term 'scribe' is and can be used. A scribe might be an official tasked with writing documents according to authoritative requirements or someone who records a document or letter based on another's dictation—whether a professional scribe, a semi-literate individual, or a member of the same household. Thus, a scribe could encompass anyone who records information. Here, the term 'scribe' is not restricted to writers of high-status texts but includes a broader spectrum.

Most contributors pay special attention to papyri and *ostraka* from Egypt due to their importance as source material, although other contexts are also considered. The data include the language use of notaries in the acts of the Council of Chalcedon and Late Latin charters from Tuscany. Cross-cultural effects on language use play a prominent role, especially the transfer of linguistic elements and scribal practices between languages.

Greek papyri from Egypt exhibit significant variation. Some variants are classified by modern editors as nonstandard, substandard, or even mistakes. Consequently, editors often 'regularise' these deviations to conform to a 'standard' Greek or Latin.[1] However, both languages underwent considerable

---

[1] Cf. J. Clackson, *Latinitas,* Ἑλληνισμός and Standard Languages, *SSL* LIII (2) 2015, pp. 309–330.

change between the third century BCE and the seventh century CE. Additionally, ancient authors were aware of sociolinguistic registers. For example:

> 1. Aristotle: οὐ γὰρ ταὐτὰ οὐδ' ὡσαύτως ἀγροῖκος ἂν καὶ πεπαιδευμένος εἴπειεν'(Rhet. 1408a).
> 'for the uneducated man will not say the same things in the same way as the educated' (my translation).

> 2. Isidorus: In quo genere dictionis illa sunt maxime cogitanda, quis loquatur et apud quem, de quo et ubi et quo tempore (Orig. 2.14.1–2).[2]
> 'In all kinds of speech, one must especially consider: who speaks, in what situation, about what, where, and when' (my translation).

When substantial variation occurs in synchronic written language, it may reflect changing attitudes towards higher linguistic varieties or shifts in education and writing skills. As literacy expanded, greater variation emerged, with many writers not mastering established spelling conventions. Different linguistic goals might also apply to different registers. It is therefore pertinent to explore what scribes deemed appropriate language for various contexts. What kinds of changes did they make to their texts, and when?

The findings reveal distinct patterns: contracts, judicial documents, letters, and petitions tend to feature more corrections to orthography and morphology, whereas receipts, administrative texts, lists, and accounts exhibit more semantic corrections. These differences align with the functions and scribal objectives of these text types.

Here is a summary of the contents of the chapters.

Sonja Dahlgren highlights that some phonological variation of Greek in Egypt most likely results from Egyptian speakers (here L1) either mishearing sounds or mispronouncing them due to limited familiarity with the foreign phonemes. Thus, various types of variation found in Greek papyrological documents may stem from a contact linguistic setting, particularly due to the influence of the

---

[2] This quite early understanding of human communication can be compared to Harold Lasswell's invention of the same idea, which became a revolutionary model of communication focusing on 'Who (says) What (to) Whom (in) What Channel (with) What Effect' (H. Lasswell, 1948: The Structure and Function of Communication in Society, in Bryson, L. (ed.) *The Communication of Ideas.* New York: Institute for Religious and Social Studies, pp. 37–51.

speakers' native language, Egyptian-Coptic (L2). The attested variation emerges through multiple language transfer phenomena, especially underdifferentiation of foreign sounds and stress-related changes. The phenomena include shifting stress positions and phoneme redistribution based on L2 stress patterns.

One notable example involves 'iotacism' (the merging of different vowel sounds into the vowel /i/), along with variation related to the rounded vowels (/o/ and /u/), which evolved due to the combined developments in Greek and the L2 Greek spoken by the Egyptians. For native Greek, this vowel variation can be tied to case merger, while in the Egyptian L2 Greek, it reflects both case merger and vowel reduction in unstressed positions.

Notably, underdifferentiation seems more linked to what speakers misheard, while stress-related variation reflects an active adaptation where Egyptian L1 speakers altered Greek sounds to fit the phonological rules of their own language.

Joanne Vera Stolk examines scribal corrections in documentary papyri, exploring their frequency, nature, and connection to the stages of document production. The text draws examples from archives, such as the letters of Apollonios and contracts from the offices of Kronion and Petaus, to illustrate these patterns. In letters, even final versions show small corrections made during composition, while non-final contracts frequently feature content-level changes in early drafts. Corrections appear more frequently in petitions and letters compared to contracts and lists reflecting differences in structure, purpose, and production processes. In petitions and lists, corrections often involve lexical and phrasal changes, which align with their complexity and the preliminary stages of their composition. By contrast, contracts typically display grapheme and morpheme corrections, especially in final versions, where semantic revisions are rare, likely due to the legal constraints on altering finalised juridical texts. Letters often combine the preliminary and final stages of composition, featuring minor adjustments to grammar, spelling, or wording upon rereading.

Documents are usually produced in preliminary and final versions, with early drafts characterised by frequent revisions, including deletions, insertions, and content changes. These documents may be created through free composition or copying, with errors and corrections arising at different points in the process. Preliminary drafts tend to include extensive lexical and phrasal corrections, reflecting efforts to refine content and formulation. Grapheme and morpheme corrections, however, are more common in final versions, often addressing minor inaccuracies or errors introduced during copying. In drafts

of lists, revisions often reflect adjustments to content rather than to linguistic formulation, further emphasising the connection between correction types and document purpose.

Marja Vierros and Erik Henriksson explored the use of computational methods for authorship attribution on Greek documentary papyri. Their approach involved marking the known authors and writers in the metadata for each act of writing, and then studying how the algorithm performed with these text parts without accessing the metadata. Despite the challenges posed by these texts, both profiling (clustering) and attribution (classification) algorithms showed promising results, though with varying success across the tested corpus from three archives dating to the Ptolemaic period. In some archives, texts clustered effectively by author, while in others, factors such as text type or professional conventions (e.g., those in notarial documents) played a larger role. The influence of writers was less significant than expected, with authorship emerging as a stronger factor in clustering. Classification results were particularly encouraging, indicating that short, fragmentary texts can be reliably attributed under certain conditions, though caution is needed when interpreting these results due to the small sample size and potential biases. While the findings support the potential of computational methods for analysing ancient texts, they also highlight that results may vary depending on the specific characteristics of the texts and archives studied. It should also be noted that the concept of 'author' in modern authorship attribution studies differs somewhat from the reality reflected in ancient historical sources. In ancient texts, the author and the writer may be distinct individuals, each influencing the text in different ways, whereas in modern studies, the author and writer are typically regarded as the same person. Nevertheless, the study by Vierros and Henriksson suggests that the writer's influence in a text does not obscure the author's role.

Carla Bruno focuses on syntactic variation within Early Ptolemaic Greek, particularly concerning the shift from the infinitive as a primary subordinating strategy to the use of finite complements (e.g., ὅτι and ἵνα/ὅπως) and the articular infinitive. Especially in Classical Greek infinitives served multiple functions, but in Modern Greek, infinitives have disappeared, leading to the development of alternative strategies for complementation. Bruno examines the transition period, particularly within a corpus of Early Ptolemaic private letters, where infinitives still appear frequently but alternate with other structures like finite complements (ὅτι, ἵνα/ὅπως), and the articular infinitive.

The analysis shows how these alternative structures began to replace infinitives, often appearing outside their original uses. For example, ἵνα/ὅπως complements start to replace prospective infinitives, reflecting preferences for clear clause boundaries and independent subjects in complements. These shifts reveal a broader trend in Greek syntax and suggest that stylistic and individual (idiolectic) factors also influenced these choices, creating a complex picture where syntactic and stylistic elements together shape the language's evolution.

Ruth Duttenhöfer's study focuses on a set of early Roman tax receipts from Elephantine, an Egyptian city located at the southern border of Egypt. The shift from Ptolemaic to Roman rule brought significant administrative changes, including the temporary disappearance and subsequent re-emergence of Demotic receipts. Under Ptolemaic rule (3rd century BCE), both Egyptian and Greek-speaking tax collectors worked together, sometimes separately and other times in collaboration. However, during the 2nd and 1st centuries BCE, Demotic receipts virtually disappear. It is only after the Roman reorganization of the Greek tax system that Demotic tax receipts reappear, particularly after the reign of Augustus.

The receipts under study, dated to the transition from Ptolemaic to Roman rule, are all traced to a single scribe, identified by his distinct orthographic, linguistic, palaeographic, and formulaic idiosyncrasies. Usually, the consistency of orthography in these receipts was strong, with few variations, except for occasional differences in the spelling of Egyptian names. These differences were likely due to scribes grappling with Greek equivalents of Egyptian names. Duttenhöfer identifies an unusual pattern in the spelling of the term *laographia*, referring to a specific type of poll tax. From the reign of Tiberius to Nero, several irregular spellings of this term are observed, including λαγραφία, λαυγραφία, λαουγραφία, λευγραφία, and λωγραφία.

Examining the published *ostraka* receipts of the early Roman period, she traced four examples of the spelling λαογραφία to a single scribe. His distinctive handwriting suggests that he was an Egyptian scribe writing in Greek, further confirming his likely bilingualism and the challenges he faced in rendering Greek terms with his Egyptian linguistic background. This highlights the complex interplay between language, administration, and identity in early Roman Egypt, shedding light on how local scribes navigated and adapted to the linguistic demands of a new ruling power.

Klaas Bentein's primary goal is twofold: first, to highlight the significance of linguistic variation in establishing social identity, particularly in documentary

texts where variation is influenced by factors such as text type, social status, and formality. Second, he argues that, beyond linguistic register, other dimensions—such as language choice, handwriting, writing material, and document format—also play a role, aligning with social semiotic concepts like multimodality. Bentein focuses on the Nepheros archive (ca. 300–400 CE), a collection of 42 texts documenting the activities of a monastic community in Egypt. The archive reveals diverse contexts, including letters, contracts, prayers, and property transactions. It shows various forms of variation: frequent orthographic errors, syntactic differences among addressors, and the use of both Greek and Coptic languages. Two document formats are noted: text written vertically and horizontally. The texts also exhibit handwriting variation, particularly in the work of certain scribes.

Tommaso Mari discusses the linguistic features of the Acts of the Council of Chalcedon (451 CE), where transcripts of debates provide valuable insights into spoken Greek in the 5th century. The theological controversies that arose within the Christian church of Late Antiquity led to the convocation of several ecumenical councils, where bishops from across the Christian world, particularly from the Greek-speaking regions, gathered to discuss matters of doctrine and church politics. The Acts of the Council of Chalcedon include correspondence and documents relevant to the council and, most interestingly, allegedly verbatim transcripts of the discussions. By comparing oral and written forms of Greek, Mari reveals how scribes normalised spoken language in their transcripts, adjusting for differences in syntax, vocabulary, and grammatical structures. He highlights the contrast between spontaneous spoken language—characterised by shorter sentences, simpler syntax, and a more limited vocabulary—and the formal written language used in the council's records.

Victoria Beatrice Fendel investigates early Byzantine Egypt, where people were immersed in a bilingual Greek-Coptic environment, and language acquisition affected language use, particularly for those who learned one language as a second language. Through analysing complementation patterns of high-frequency verbs (γράφω, οἶδα, and θαυμάζω), Fendel explores how scribes learned Greek. Based on a corpus of 264 bilingual letters, she examines linguistic variation due to register, internal confusion, and bilingual interference from Coptic. Different learning approaches—pattern-based, exemplar-based, and chunking—are explored in relation to the use of standard collocations, idioms, and formulae for these verbs. The results suggest that Greek writers often let regular patterns intervene in formulaic contexts, replacing idiomatic uses with regular forms, and

occasionally relied on Coptic structures. In contrast, Coptic writers consistently used synchronic standards, suggesting less syntactic ambiguity. This discrepancy hints that pattern-based and chunking approaches were integral to Greek learning, with collocation confusion signalling an early learning stage and idiom confusion indicating a more advanced one. The study concludes that Greek-Coptic bilingualism and different learning strategies significantly impacted scribes' linguistic choices.

Tonio Sebastian Richter discusses the significance and unique features of an exceptional papyrus, the *P. Budge* hearing protocol, an ancient document that provides valuable insights into scribal practices, language use, and judicial customs of the time. The *P. Budge* is an extraordinary document that was preserved possibly due to the personal connections or status of the parties involved in the hearing protocol copied on the papyrus. The document is, therefore, not the original transcript from the hearing, but a clean, corrected copy signed by participants. Its linguistic profile reflects the Sahidic dialect of Coptic with nonstandard features, some influenced by Upper and Middle Egyptian. The text also incorporates an unusual number of Greek words, including rare juridical terms and function words, some of which are typical of one of the persons involved, Philemon.

    While most of the hearing's content seems to have been orally performed, parts of Philemon's speech may have been written in advance and later integrated into the protocol, as written elements, such as phrasing and formatting typical of petitionary letters, support this conclusion. The hearing also reveals important social implications. The advanced rhetorical style and innovative language of Philemon, a peasant, suggest external assistance from experienced advisors, raising questions about his social status and connections. His ability to maintain this document in a private archive hints at his economic power and unique circumstances. *P. Budge* offers remarkable linguistic and historical insights, blending individual expression with broader judicial and scribal traditions, though questions about its preservation and Philemon's status remain unanswered.

Timo Korkiakangas explores Early Medieval Latin documentary texts from Tuscany in 8th- and 9th-century Latin charters. He analyses the potential relationship between spelling variation and the use of specific grammatical forms. Using the Late Latin Charter Treebank (LLCT)—a corpus of 200,000 words from 8th- and 9th-century Tuscany—Korkiakangas analyses 519 charters by 176 scribes. Charter Latin, while often nonstandard, is compared to Classical Latin norms, to which scribes still referred to varying degrees.

The hypothesis suggests that scribes who used more nonstandard spellings were also more likely to produce innovative Romance-like forms and constructions and struggled more with Classical Latin forms that were fading in Late Latin. These innovative forms entered writing from spoken language, while conservative forms originated from traditional legal Latin, memorised and reproduced with mixed accuracy. Korkiakangas uses a generalised linear model (GLM) to analyse how spelling variation correlates with 11 linguistic features associated with language change and the year of writing. Standard Latin spelling correlates significantly with lexical and morphological features, but less so with syntactic features. This distinction between syntax and other grammatical areas likely reflects the way Latin was taught, with syntax being less successfully acquired by L2 learners compared to less abstract areas of grammar. Thus, even skilled spellers used Classical Latin syntax inconsistently. The study also acknowledges that other linguistic and extra-linguistic factors, such as document formulaicity and individual preferences, could influence these findings. Spelling choices correlate with both conservative and innovative features of Latin.

The basic analysis of the chapters can be summarised as follows:

1. Who produced the language?
A prototypical scribal text is one in which the source of the content differs from the person who pens the text. But how can we determine which parts of the language, if any, derive directly from the author? Can modern authorship attribution methods help to provide an answer? In letters, it is natural to assume that most of what was written originated in one form or another from the author. However, even in documentary texts, there are often passages where the commissioner had to be active in formulating the content. In other instances, we know that the scribe played an active role in producing and modifying the text—for example, by re-writing notes made from an oral statement into written form or by creating a text for professional purposes. In these contexts, it is also necessary to examine the correlation between different documents, or parts thereof, and their linguistic features.

2. Scribal choices and corrections
The activity of scribes is evident on various levels of text production: orthography, syntax, palaeography, language choice, document format, and layout. After composing the text, a scribe may have wished to insert corrections. But what did the scribes correct, and where? Scribal texts may display considerable variation

across linguistic levels, and it is intriguing to see whether nonstandard forms on one level (e.g., orthography) correlate with those on another level (e.g., syntax). The question of different factors causing linguistic variation may also be explored in an archival context, where the researcher is usually better informed about the context than in the case of documents preserved in isolation. Scribal activity naturally encompasses the palaeographical level. Many, if not most, scribes in late antique Egypt were accustomed to writing both Greek and Coptic (and even Latin) documents. In such a context, it is conceivable that, despite the differences in alphabets, certain convergences developed in individual letter forms.

3. Language-internal pressure

The scribe may have been unaware of or uncertain about the correct linguistic form. This could have been due to imperfect knowledge of the correct phraseology or to pressure from (e.g., phonological) variation in the language used, leading to a clash between a linguistically motivated form and a standard form. Language-internal variation and change, together with the scribe's competence, interact and shape the scribes' choices. This may result in considerable variation, particularly in contexts where extensive reorganisation of the linguistic system was underway. The spoken level of the language may become visible in the written version in various ways.

4. Language-external pressure

The clash between different forms and the resulting uncertainty may also have arisen from language-external pressures. The scribe may have been writing in a language that was not his L1 or, in any case, was not the language in which he was most literate. In a society where more than one language and script were used to convey information, scribal choices concerning language use, and the presence of multiple languages and scripts within a single text, can be revealing. A scribe might also be a language learner, and in these contexts, we observe various degrees of linguistic transfer from the scribe's L1 to L2. Scribes tried to produce language that aligned with his concept of correctness and adhered to the conventions of each genre (such as poll tax receipts and private letters). In this respect, we propose extending the somewhat problematic term 'scribe' to include individuals who wrote, for instance, letters for themselves. In this sense, a scribe is simply a person who writes. This broader definition allows us to include private letters—a corpus that is highly rewarding for any linguistic research conducted on papyri.

Martti Leiwo

1. The Language Use of the Narmouthis Scribes: Foreign Language Perception and Native Language Transfer. A Case Study

Sonja Dahlgren

1 Introduction[1]

In this Chapter, I explain the background theoretical reasons for some of the variation in Greek used in Egypt, as it resulted from the language contact between two very different types of languages. I will show examples of nonstandard language usage that I believe to be the product of the language contact between Greek and Egyptian-Coptic in (mostly) Roman period Egypt, concentrating on the phonological level. Based on the evidence of nonstandard Greek used in Egypt, when compared to that in the 'mainland' Greek, it seems reasonable to treat it as a contact variety of its own (see more Dahlgren 2022 and Dahlgren et al. 2022 with relevant references). This conclusion is largely based on the phonetically-based misspellings in the documentary material, which, in a phonological analysis, revealed contact-influenced reasons for some of the nonstandard vowel orthography (see Dahlgren 2017 for a comparison to Coptic phonology; see also Dahlgren (accepted) for a new analysis of iotacism in Egypt). There were two different avenues working together in how this variation might have evolved from a second language (L2) version of Greek to an independent Greek variety: foreign language perception and native language transfer. I will look at how both hearing and producing a foreign language can affect its treatment in a new linguistic environment, as separate routes to new variation.

The Narmouthis Greek ostraca, the focal point of the phonological analysis in this Chapter, are suitable for a representation of the contact phonological phenomena occurring in Greek in Egypt for many reasons. First, the collection represents one of the earliest examples of Roman period documentary text

material written in Greek (1st – 3rd century CE).[2] Second, it is from Fayyum, one of the most bilingual areas of Roman Egypt. Third, the collection is bilingual with Greek only, Demotic only and Greek-Demotic texts, a testimony in itself of a contact linguistic scenario. Fourth, some of the Demotic texts contain elements of a very primeval form of (pre-Old-)Coptic (e.g. Quaegebeur (1991: 190–1); Grossman and Richter (2017: 215 fn9, 231); summary in Dahlgren 2017: 31–4). This allows a comparison of the Greek phonetically-based misspellings with similar ones from the Coptic documentary genre. Finally, the original purpose of the texts has been much speculated and often considered to belong to a school milieu (Bagnall 2007: 21; Fewster 2002: 235). The fact that the writers might have been pupils or scribal apprentices gives a direct view into the linguistic background of the writers: if, due to unfinished Greek education, the standard orthographic forms of words were not always remembered or were misremembered, misspellings based on the phonetic forms of the words were used in their stead. This last aspect gives wonderful opportunities for contact phonological analyses of the language use of these writers, depending on the type of the misspellings. Using phonetic forms of the words proves that the scribes knew how to speak Greek; similarly, as some misspellings present phonological and phonetic features that were not part of the Greek-internal developments, these spellings also give valuable information of Coptic phonology.

This paper is organised as follows. After this general introduction to the topic at hand, I will give a very brief general introduction to the stage of the Greek phonological development and how it was realised in Egypt in Section 2. In Section 3, I will discuss the theory of language contact emerging through inaccurate listener perception of a foreign language. In Section 4, I will present the Egyptian Greek variation in the context of listener- vs. speaker-induced sound change. In Section 5, I discuss the main results of this study.

---

[2] The Narmouthis Greek ostraca (OGN I; Pintaudi & Sijpestein 1993); in this paper, the Narmouthis Greek Ostraca are referred to as O.Narm. as in The Papyrological Navigator (papyri. info), which lists all the texts in OGN I. While the examples have mainly been taken from the Narmouthis collection, some of the same type of variation has also been found in the Eastern Desert text collections, military camps, as, for example, described in Leiwo (2010) and (2022). They do not display all of the vowel variation types that are present in Narmouthis Greek, however, that directly compare with Coptic vowel inventory; for instance, lacking the variation related to /u, y/. Besides this detail, they are extremely important for the study of Egyptian Greek as a contact variety, as they show the spreading of some of the Egyptian-Coptic-induced vowel variation across Egypt (see further in Dahlgren 2016, 2022 and Dahlgren et al. 2022).

2 Greek phonological development and features of Greek in Egypt

Greek went through some major phonological changes from Roman period onward that resulted in many of the vowels raising and fronting to /i/ (e.g. Horrocks 2010). However, two things should be considered when looking at Greek orthographic variation during this period: first, there is an abundance of phonological variation that seems unusual for native language (L1) Greek users, and second, most of the Greek material from the first centuries CE comes from Egypt, which had Greek as the official language of the government. Language contact, therefore, should be considered to be the reason behind some of the more creative nonstandard spellings; indeed, some of the variation presents obvious confusion or merger of phoneme qualities that are distinct phonemes in Greek, such as between /o/ and /e/ (Dahlgren 2017: 62–5). Comparing these nonstandard spellings to evidence known of Coptic phonology (e.g. Kahle 1956, Peust 1999), patterns of transfer of Egyptian-Coptic phonological qualities on the L2 Greek used in Egypt start to emerge. The main, most frequently appearing features typical for L2 Greek in Egypt have been presented in Gignac (1976), Teodorsson (1977), Horrocks (2010) and Torallas Tovar (2010). The most notable changes that developed in the Greek vowel inventory were, of course, the qualities of those vowels that were included in iotacism and gradually all merged with /i/: the monophthongs /eː/ and /y/, and the two diphthongs /ei/ and /oi/. Furthermore, the voiced plosive series, as well as the last elements of the diphthongs /eu/ and /au/, developed into fricatives. In Egyptian Greek, however, the variation is in part very different. The voiced plosives are often replaced with voiceless ones, and vowel variation seems to come from another planet. These are all features that derive from the contact with Egyptian-Coptic, which had no opposition between voiced and voiceless stops, and whose vowel system (if not inventory) was quite significantly different from the Greek one. We know vowel qualities from Coptic, and we therefore know that while Coptic had no /y/, as Greek did, it did have most of the other vowels in Greek vowel inventory: /i, e, a, o, u/, and in addition, /ɛ/ (Peust 1999: 201). But how these vowels were used differed from the Greek usage, as Coptic had vowel reduction, while Greek did not.

It seems, therefore, clear that much of the nonstandard variation in L2 Greek documentary texts is the result of this language contact situation, and direct transfer from Egyptian-Coptic. But how exactly does this happen, in the language use of an individual speaker? Do we perceive the foreign sounds incorrectly, or just fail to produce them adequately? And how could this be

known of speakers long gone, examining their language use in historical written material? This will be discussed in Section 4, but let us first look at some of the contact-linguistic phonetic phenomena that lead to this language variation and change.

3 Language change caused by listener perception[3]

There is nothing particularly strange in the Greek-Egyptian language contact that would not have been described by Haugen (1950), Weinreich (1953), Thomason & Kaufman (1992), Thomason (2001), Sankoff (2004), Matras (2009) or numerous other contact linguists. In fact, the existing theories regarding the type of contact features and level of integration of the L2 Greek speakers in relation to the motivational status of the language learners (especially regarding professional scribes), as well as the type of contact regarding its intensity, match the hitherto descriptions very well (see e.g. Matras 2009: 223–6; related to Egyptian scribes, see Dahlgren 2017: 73–4).[4] Therefore, instead of proving the nonstandard features were contact-induced, in this paper I concentrate on the *why* instead of the *what* in terms of explaining the phonological variation during the language contact that formed what seems to have been a contact variety of Greek in Egypt (Dahlgren 2022).

It is usually assumed that contact features of e.g. underdifferentiation, phoneme replacement and stress transfer are formed through imperfect L2 speaker production (e.g. Thomason 2001: 75, see also Ohala 1981: 178). It is also understood that L2 production of this type is most often affected by L1 phonology and what is available in it with which to represent the foreign phonemes (Haugen 1950; Weinreich 1953 etc.). However, as Haugen pointed out (1950: 216–7), the very beginning of language contact is usually started by a small group of bilingual speakers who spread the use of the first loanwords into the monolingual group. The mimicking of these foreign words by the monolinguals at first shows extensive phoneme substitution and irregular production of the target phonemes, which causes fluctuation in the pronunciation of the foreign

---

[3] Many thanks to Pertti Palo for invaluable comments on the theoretical description of acoustic phonetics.

[4] See also e.g. Vierros (2012) related to morphosyntactic transfer phenomena in the language use of Upper Egyptian scribes before the Roman period, testifying to a wider contact linguistic situation between Greek and Egyptian.

phonemes. This variation in the production of the foreign words is further copied by other monolingual speakers and spread as a distorted version of the original sound qualities. Therefore, the starting point of language change is the hearing of a distorted version of the L2 words, biased by L1 phonology, and the repetition of this nonstandard production to others, as described by Ohala (1981). It is only after this initial step that on further repetition of this changed version of the L2 phonemes, contact varieties are born (Thomason 2001: 75 on Indian English).

However, fluent bilinguals may also converge the phonological systems of the two languages they speak so that the same phoneme inventory is used for both languages, as in most of the Anatolian Greek dialects' replacement of the inherited dental fricatives /θ, ð/ by /t, χ/ or /d, r/ (alternation dialect-dependently) because the Greek original dental fricatives do not exist in the other language in the contact situation, Turkish (Matras 2009: 229–30; Zimmer and Orgun 1992 on Turkish consonant inventory). In any case, variations of phonological production that are somehow influenced by the native language, or simply the stronger of the two (or more) regularly used, contribute to the emergence of contact varieties. Essentially, there is a combination of both hearing and reproducing the foreign elements in an altered manner, which, obviously, is further challenged if the languages are typologically dissimilar. Haugen gives an example of a contact situation between Yaqui and Spanish. The Spanish loanword *estufa* was pronounced [ehtúpa] by the first-generation Yaqui-Spanish bilinguals, the L2 speakers thus having replaced Spanish /s/ with /h/ and /f/ with /p/ under the effect of the L1 phoneme inventory (Haugen 1950: 217) – in other words, a change in the place of articulation of the fricatives from alveolar to glottal, and a change in the manner as well as place of articulation in the labials, from a labiodental fricative to a bilabial stop. In his article, Ohala (1981) studied the listener's role in sound changes that concern especially contact situations between unrelated languages, as these often have a phonetic origin. Such unrelated languages are, for example, Greek and Egyptian-Coptic, providing an interesting case study when looking at the different origins of contact-induced variants; some nonstandard features are a result of inaccurate production, while some derive from imperfect hearing.

Until Ohala's (1981) paper on the origins of sound change, it had generally been assumed that effects of language contact on the pronunciation of a foreign language derived from compromised production. Most former theories were speaker-oriented: that the speaker modified their pronunciation

to reduce the energy used in speaking, or made their speech more distinct to enhance intelligibility, or to simplify grammar (Jespersen 1949: 15ff; Martinet 1964: 169ff; Halle 1962 (as cited in Ohala 1981: 178)). While all that is undeniably true, there is something that happens in addition to the production of a foreign word, and that is, of course, the perception of it by others. Ohala pointed out that when less fluent bilinguals or complete monolinguals hear a foreign word, no matter how perfectly pronounced by the L1 speakers of the language or fluent bilinguals, the pronunciation of it might get distorted when again produced (see also Haugen 1950 on this). The type of language change following from this type of variation excludes language-specific and culture-specific sound changes that may have an origin in e.g. paradigm regularisation. The foundation of Ohala's theory lies in the simple fact that the acoustic speech signal is inherently ambiguous due to the many-to-one relationship between vocal tract shape and sound shape, which might affect articulation of the sound. Therefore, what one speaker produces may not always be successfully repeated by another one, who, trying to resolve this ambiguity, produces a different articulation from other speakers. This results in sound changes such as the English word *with* being pronounced [wif] from the original [wiθ] in some English (mostly working class) dialects such as Cockney, Essex dialect, Estuary English, some West Country dialects, Yorkshire, and even some Scottish English varieties (although partly conditioned by specific contexts) and e.g. African American vernacular English and Liberian English (see *th*-fronting, e.g. Wells 1982; Tollfree 1999; Schleef and Ramsammy 2013). Similarly, the Proto-Indo-European *$*g^{w}\bar{o}ws$* 'cow' was realised as *bous* in Greek, replacing one manner and place of stop with another. These variants are different in terms of articulation but very similar auditorily – it is therefore an easy mistake to make when hearing something imperfectly and repeating the sound as accurately as possible, i.e. phonetically (Ohala 1981: 178).

Another example, perhaps even more fitting for the Greek-Egyptian contact, concerns the acoustic consequences of consonant and vowel interaction, i.e. their coarticulation. Figure 1 from Ohala (1981: 180) shows how sounds can become distorted in language contact.
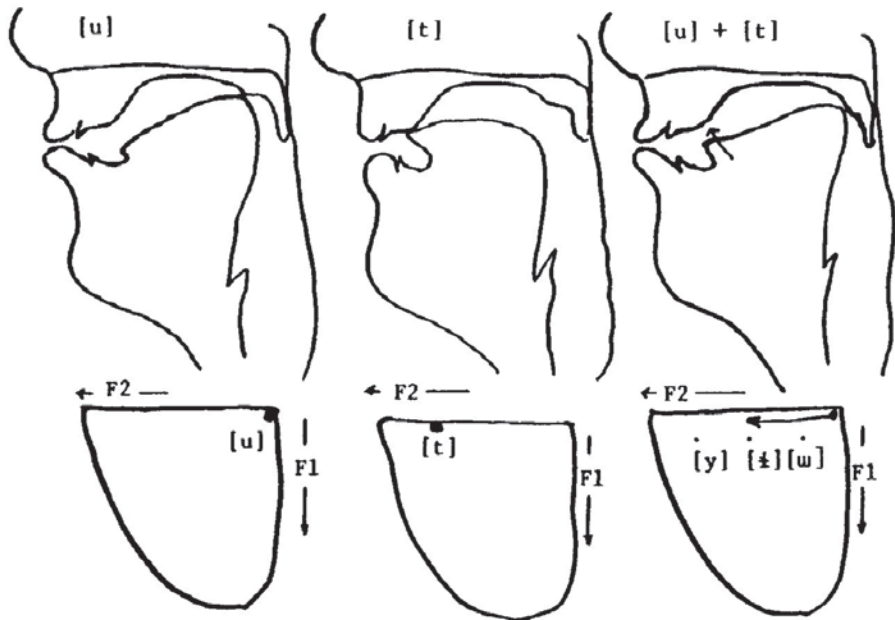
Figure 1. Vocal tract shape (top) and acoustic output (bottom) of [u], [t] and simultaneous coarticulation of [u] + [t].

Figure 1 The production of /u/ and /t/ in coarticulation (Ohala 1981: 180).

The vocal tract shape when uttering /u/ produces a sound with a low formant 1 (F1) and formant 2 (F2).[5] When producing a dental/alveolar stop /t/ after /u/, however, the vocal tract is more constricted at the front than usually for /u/, which creates a sound with a low F1 but a higher F2 (ca. 1800 Hz – in the normal (male) production of /u/, the average height of F2 is about 595 Hz; Catford 1988: 161). If these sounds are coarticulated i.e. produced in rapid sequence, almost simultaneously, (etc.), when they can affect the pronunciation of one another, F2 is determined by the apical constriction because tongue is moving higher and frontward to target the production of the dental/alveolar stop /t/. This means that F2 is also higher when producing /u/, which would need a lower F2 to sound like the back vowel it is. Consequently, the production of /u/ is fronted in the acoustic/auditory space, and is phonetically nearer to [y], on a scale (back

---

[5] Vowel quality can be measured from a spectrogram as the time-varying peaks we call formants. For vowel qualities, especially the first formant, F1, and the second formant, F2, are important. F1 corresponds to vowel openness and F2 to its frontness. Open (or low) vowels have high F1 frequencies and close (or high) vowels have low ones; back vowels have low F2 frequencies and front vowels have high ones (see e.g. Reetz & Jongman 2009: 182–4).

to front) from [ɯ] to [ɨ] to [y], depending on individual variation. This concerns any (rounded) back vowels in the environments of consonants with a high F2. Naturally, the listener is faced with some challenges when trying to repeat what the speaker said, due to that above and other vocal tract constraints that distort the speech signal, resulting in productions not always intended by the speaker. Essentially, the listener of this type of a distorted production will have to decide which phonetic events in the acoustic signal they should try to actively repeat and which ones to ignore as meaningless variation caused by the mechanical properties of the vocal tract. In the production of the sequence /u/+/t/, the listener would have to learn that the vowel target is /u/ and not the variants [ɯ], [ɨ] or [y], i.e. they would have to learn to disregard the pitch difference after the stop. Normally, the listener does learn these details of speech production, within years of practice of hearing and repeating the sounds of their (native) language. But if the listener is trying to repeat sounds that are foreign to them, the normal rules they have learnt suddenly do not apply: e.g. (American) English speakers may not produce the Arabic sound [ʕ] in the right way, including the tense voice quality that accompanies the production of the pharyngeal sound (Ohala 1981: 180–181).

Foreign language pronouncing is challenging, which is reflected in the vast amount of literature on second language speech perception studies (e.g. Strange et al. (2008); Strange and Shafer 1995; Major 2001). There are two basic hypotheses. Previously, it was believed that second language learners mainly operated by replacing the foreign L2 sounds by similar sounds available in their own language, which also means that they could not distinguish them when hearing them (see e.g. Weinreich 1953; Strange 1995). According to this hypothesis, all foreign sounds, whether only slightly different or more so, would have been filtered through the L1 phonological system. However, more recent studies (Major 2001: 37; Strange et al. 2008) reveal that L2 learners might even learn completely novel foreign sounds better than those that differ only slightly from the phonemes in their own language, i.e. phonemes that are similar but acoustically non-identical to those existing in the L1 phoneme inventory. They may remain imperfectly learned as the subtle differences may not be auditorily clear to the L2 learner. The more noticeable differences, however, are perceptually more salient, so the language learners' attention focuses on them, which, consequently, results in learning them better. For instance, according to Major (2001), an L1 English speaker of L2 French might substitute the French unaspirated dental /t̪/ with the English aspirated alveolar /tʰ/ because the difference is very small. However, as Major continues, when the L2 phoneme is foreign enough, i.e. there is no

existing model even nearly like it in the L1, there is nothing to transfer from L1 to L2 anymore. Then, the L2 learner makes an attempt at learning the new phoneme, as in for instance managing to learn the uvular rhotic sound /R/ of French, even though the rhotic sound used in the L1, English, is very different (Major 2001: 37). At least for the latter example, this is no doubt a more speaker- than listener-oriented task, even at the stage when the rhotic is still imperfectly learned and substituted with the L1 rhotic (a (voiced) postalveolar approximant /ɹ/ in Standard British English). At this stage, the L2 learner can perhaps hear the difference but has not yet learned to produce it; the basic sound substitution as already described by Weinreich 1968: 18–19).

Both of these situations are relevant for the Egyptian L2 speaker of Greek. As is clear from the misspellings in Egyptian Greek texts, the voiced plosives γ, δ /g, d/ and (to a lesser extent) β /b/ were often misspelled as κ (1a), τ (1b), π (1c).

> (1a) γεορκους < γεωργοὺς (O.Narm., 8; /g/ → /k/)
> (1b) ατελφης < ἀδελφῆς (O.Narm., 28; /d/ → /t/)
> (1c) Κλευπις < Κλέοβις (O.Narm., 22 & 25; /b/ → /p/; Dahlgren 2017: 84 ff.)[6]

These nonstandard orthographic variants resulted, naturally, from the altered pronunciation of these sounds by the Egyptian writers: these phonemes did not exist in the Egyptian-Coptic phoneme inventory, so they were underdifferentiated and substituted with the closest L1 equivalent sounds (see Weinreich 1968: 18–19). The change is not dramatic: it concerns lack of voicing where it should be produced, but the sounds are still plosives, and words mispronounced this way would have been understandable as substitutes even to the L1 Greeks. Regarding the vowels, there was much more variation that could have led to confusion of lexical content as they were, for instance, responsible for case marking. Greek /y/ represents the first situation described by Ohala (1981), a foreign phoneme not existent in Egyptian-Coptic; regarding the several vowels connected to iotacism, the situation is similar to the situation referring to the replacement of the French dental stop with the English alveolar one, as described by Major above. It seems probable based on the many misspellings that when the Egyptians were faced with the sudden need to deal with the Greek language, they had little time to learn it. Consequently, the long time of practice mentioned by Ohala to learn the phonemes as they were meant to be pronounced in native Greek was not available,

---

[6] Many of the examples given in this paper have been presented in Dahlgren (2016) and (2017) with more detailed phonological analyses of the nonstandard variation.

and especially the first generation of L2 Greek speakers, therefore, understandably made errors when trying to repeat the foreign sounds. This matches the description of Haugen's (1950) 'pre-bilingual period' (Haugen 1950: 216–217): the first small group of bilingual speakers uses loanwords from a foreign language, and these are distributed to the monolingual speakers who fail to pronounce them accurately. The following generations, theoretically, learned the phonemes better; it is true that the most notable 'Egyptianisms' in the L2 Greek production are from the early Roman period centuries, in Greek as well as Greek-Coptic bilingual texts (see e.g. the many similar misspellings in the manuscript P. Hamb.bil. 1; examples in Dahlgren 2017). This would have co-occurred with a sudden vast increase in the amount of scribes writing in Greek for the Roman government, and the emergence of Coptic (see a more detailed discussion on this in Dahlgren 2022: 115–118). But to some extent, there must have been fossilisation of some Egyptian-Coptic -influenced phonemes that ended as part of the Egyptian Greek variety (see Dahlgren 2022, Dahlgren et al. 2022), as even very late native Coptic texts still have some of the same nonstandard variation that is seen in the earliest Roman period texts: variation between voiced and voiceless stops, replacing *eta* with *epsilon*, etc. However, the variation between /y/ and /u/ disappears after a few early Roman period centuries, probably indicating its gradual loss in Greek itself, whereupon the new generation of L2 listeners would learn to produce a different vowel (one closer to [i]) when trying to repeat it. Be that as it may, there are two phenomena that can be noticed in Egyptian Greek: underdifferentiation of foreign sounds, being replaced with the nearest native language equivalent (voiced and voiceless stops, /y/ with /u/, much of what is considered iotacism), and more 'active' transfer of L1 phonological features onto the L2, such as transfer of stress patterns. The first is likely more connected to under-distinguishing these sounds auditorily, i.e. would be listener-induced perceptually-based variation. The second is connected to the replacement of the foreign sounds and stress patterns with native ones, i.e. integration, and this, I believe, would be speaker-induced variation.

The first instance is probably related to the listener not recognising the target /y/ within the phonetic 'noise' due to the phoneme being foreign, so it gets produced as the nearest phoneme available in the native language phoneme inventory. Egyptian-Coptic had /u/, so that was used, apparently as long as /y/ was used in Greek; the Egyptian L2 user of Greek seems to have recognised the roundedness in the vowel quality, even if it was too far back from the intended target (on Coptic vowel qualities in more detail, see Peust 1991: 201). When /y/ merged with /i/ in Greek, the situation changed, and /y/ was, on the grapheme level, more and more nonstandardly replaced with other sound-letter pairs

involved in iotacism: ι [i], ει [i], η [ẹ], οι [ø/y] (which, also, slowly merged with /i/ (Horrocks 2010: 167–168; on the variation between /y/ and /u/ in Egyptian Greek, see Dahlgren 2017: 68–82)). Iotacism, on the other hand, connects to such fine-grained distinctions as e.g. between Roman period *eta* pronounced as [ẹ] and *iota* as [i] that as they were frequently confused even by L1 Greek writers in mainland Greek inscriptions, they were probably not distinguishable to a listener with only one of these phonemes (/i/) in their native language phoneme inventory, and consequently, the phonemes were underdifferentiated and Greek *eta* /ẹ/ often replaced with *iota* /i/ in the texts (on vowel qualities in Greek in Egypt in more detail, see Horrocks 2010: 165–170). But related to this variation, something else happened, too. The Egyptian L2 Greek users did not merely fail to distinguish between the two (or diachronically, more than two) vowel qualities that in the end merged with /i/ in Greek: they also transferred properties of their own native language on the phonemes, by consonant-to-vowel coarticulation. To an extent, this concerned all vowel production of Greek, but iotacism was more vulnerable to influence from Egyptian-Coptic because there seems to have been a clear need for distinguishing between high and low vowels in Coptic, in order for this to aid with the recognition of the consonants surrounding the vowels (see more in Dahlgren 2020, Dahlgren accepted). Yet, the vowel variation related to iotacism is still within the realm of underdifferentiation (see Weinreich 1968: 18). More active transference of native language phonology occurred with the transfer of stress, and the phoneme distribution related to that.

4 Egyptian-Coptic -influenced phonological features in listener/speaker-oriented contact variationist context

The Greek-Egyptian language contact and its consequences for the phonology of Greek spoken in Egypt have been noted before by Girgis (1966), Gignac (1976 & 1991), Horrocks (2010: 112) and Dahlgren (2016), (2017). However, while at least part of this variation has been labelled as deriving from a language contact scenario before, it has not been considered before according to which phonetic/phonological mechanism(s) the different types of variation were born.

    In my opinion, the vowel variation in Egyptian Greek can be divided into two rough categories: underdifferentiation of foreign phonemes and stress-related redistribution of phonemes. Both are, of course, transfer phenomena from L1 Egyptian-Coptic to L2 Greek, but it could be argued that different L2 distinctions that do not occur in the L1 of the language users are underdifferentiated not

only due to the speakers not being able to produce the new phonemes, but because it would probably also be difficult to distinguish them in hearing. On the other hand, stress differences can be heard but not repeated – according to Matras (2009: 69–70), adult L2 learners have particularly problems with leaving behind their L1 prosody, it being one of the most prevalent factors behind what is considered a foreign accent (see more in Section 4.2). In Egyptian-Greek, the stress transfer from Egyptian-Coptic brought with it vowel reduction rules: based on the L2 Greek usage of Egyptian writers, it seems that /o, ɔ/ could not occur in an unstressed position in Egyptian-Coptic, and furthermore, especially word-finally, /a, e, o/ were reduced to schwa (Dahlgren 2017: 83–90, 62–65: Dahlgren 2020). These features had an auditory basis in the Egyptian-Coptic language: the positions mentioned were particularly poor for clearly distinguishing mid vowels, and these were therefore reduced in quality. Reduction to schwa word-finally was probably not observed and regulated by the speakers themselves but rather a phonetic result from voice quality weakening towards the last syllable of the word. However, regarding the stressed rounded vowels, raising /o/ to /u/, one of the corner vowels, was probably a phonological rule: to have a rounded stressed vowel whose quality would be easy to hear. The problem in Egyptian Greek was the transfer of these vowel distribution rules on Greek, which did not have the same system of vowel reduction and relied on many of the word-final vowel qualities for grammatical information, such as case or voice marking, which was lost with the contact-induced misspellings (see Dahlgren and Leiwo 2020). Next, we will take a look at these contact phonological phenomena in more detail.

4.1 Underdifferentiation of phonemes

Egyptian-Coptic phonological influence can be seen in the underdifferentiation of some Greek phonemes: all the features of the Greek phonemes were not realised in their full form but modified to fit the phonological system of the speakers' L1, Egyptian-Coptic. For instance, in O.Narm. 42, a (possible) tax receipt dealing with a payment in wheat, the Egyptian-Coptic influenced confusion of /y/ and /u/ has resulted in the under differentiation of Greek /y/ in the relevant word, πυροῦ, which has been replaced with /u/, a close-enough rounded vowel that was available in Egyptian-Coptic. The resulting misspelling was πουρου (1).

  (2) πουρου < πυροῦ 'of wheat (gen.)' (O.Narm., 42; Dahlgren 2017: 69 ff.)

The cause of variation here is the simple fact that this vowel distinction did not exist in Egyptian-Coptic (e.g. Peust 1999: 201) and therefore probably was not auditorily clear for an L1 Egyptian speaker – a classic case of underdifferentiation as defined by Weinreich (1968). The high front rounded vowel /y/ has the distinctive features [+high], [+front], and [+round] (see e.g. Jakobson and Halle 1971 for more). This foreign sound is replaced in some Egyptian Greek texts with /u/, which only has the features [+high] and [+round] of the intended phoneme, and the feature [+front] has been replaced with the feature [+back]. The replacement is as close a match as possible from the L1 phoneme inventory, an equally high and rounded vowel but back, not front. This is linguistically interesting because the Egyptian L2 Greek speakers could also have chosen /i/ as the replacement phoneme, with the features [+high], [+front] but [-round].

The high front rounded vowel /y/ is cross-linguistically rather rare; only ca. 6.6% of the world's languages have it (Maddieson 2013). Therefore, as it does not exist in many languages, it has different representations in the speech of L2 speakers of the languages that do have it: it varies between the two possibilities for it, /i/ and /u/. Both are high vowels but the first is unrounded while still front, and the second is rounded but back. For instance, for the French /y/, Arabic speakers used /i/ and e.g. Italian, Spanish and English speakers a vowel quality nearer to /u/ (Vieru-Dimulescu and de Mareüil 2006: 441, 443–4). The choice for the replacement vowel probably depends on which one is the more important distinction for the L1 of the speaker, and more generally, what is available for an L1-based substitution. In Maghreb Arabic, there are no rounded vowels so the choice is /i/; the speakers of the Indo-European languages act differently because the languages have rounded vowels. In the case of the Egyptian scribes, the distinguishing factor in terms of underdifferentiation might have favoured /u/ for its roundedness because of the tendency of Egyptian-Coptic for consonant-to-vowel coarticulation especially regarding front vowels (see more Dahlgren 2017: 74–80 and Dahlgren accepted). Thus, /u/ was a clearer representation for /y/ that still displayed roundedness. The variation, therefore, also includes useful information for dating the gradual loss of rounding in the phoneme written with *ypsilon*. Tied to the general Greek development of iotacism, the rounded quality /y/ belongs to the earlier centuries CE, even though the unrounding of it was fully finalised only by the 9th–10th century CE in Egyptian Koine (Horrocks 2010: 166-169; see also Teodorsson (1977) and Allen (1968) for general developments). However, as Egyptian-Coptic had fewer front vowels in general, and no /y/ to add to that, iotacism was, according to Gignac (1991: 187), accelerated in Egypt. This shows the role of the listener for sound change, as in the reproductions of

the Greek *ypsilon* in Egyptian Greek texts, the graphemic interchange between *ypsilon* and the diphthong ου (<ou> /u/) dates to the early Roman centuries, and only a bit later the letters used in the nonstandard variation slowly change to *ypsilon* between *iota*, *eta* etc. with the apparent contact-induced gradual loss of the roundedness of the vowel (Dahlgren 2017: 81–2). This, again, deepens the impression that the L2 Greek speakers from Egypt heard the original production well enough.

However, some of the misspellings also display an evident endeavour toward consonantal coarticulation on vowels, resulting in vowel qualities being altered as they were affected by the surrounding consonants. This was particularly frequent regarding the non-back vowels ι η ε α /i ẹ e a/, which sometimes makes it look like iotacism (Dahlgren accepted). The Afroasiatic languages have word formation based on a consonant root so distinguishing between e.g. the place of articulation of consonants is important. For example, this could mean clarifying the distinction between a velar stop and a uvular one through the changed quality of the vowel (fronting it from [a] to e.g. [e] near the former and retracting it from [a] to [ɑ] near the latter; see e.g. Bellem (2007: 174–5) on a description of the Arabic phonological system). This kind of differentiation of consonants through the surrounding vowel qualities is not relevant for Greek; in Egyptian Greek, it is merely an accidental effect of the phonological transfer from Egyptian-Coptic to Greek. In Weinreich's terms, it is overdifferentiation, i.e. transferring a feature that is linguistically significant for the first language but redundant in the second (Weinreich 1979: 18–9). It is one of the clearest examples that fit into Ohala's theory of listener-induced sound change: there is little doubt that the Egyptian users of L2 Greek heard the consonantal effects correctly, but for the sake of Greek, their hearing being affected by their L1 phonological system, they heard too much, and repeated this in their language production. Examples (3–4) show this phenomenon, with (3) offering a fronted vowel quality in the proximity of a front consonant (ρ /r/) and (4) showing a retracted vowel quality adjacent to the bilabial nasal (μ /m/), which often retracts the quality of the nearby high vowels (Flemming 2009: 82–84; 92; see Dahlgren 2020: 219–0 for more analysis). Both examples involve variation between η and ε because all other variation regarding *eta* can easily only be seen as part of iotacism (especially variation between *eta* and *iota*).

(3) Ἥρμα < Ἑρμᾶς (O.Narm., 67; Dahlgren 2017: 105–6)

(4) μετροπολι < μητροπόλει (O.Narm., 110; Dahlgren 2017: 105)

In (3) Ἥρμα, a nonstandard spelling of Ἑρμᾶς (a Doric/Aeolic dialectal form of Ἑρμῆς), we see the standard *epsilon* being replaced with *eta* in front of /r/, which is a coronal consonant. Coronal consonants are produced in the front part of the oral cavity and may thus front nearby vowel qualities, which could have happened here.[7] In (4) μετροπολι, the anticipation for the bilabial /m/ may well be behind the retracted quality of *eta* to *epsilon*.

There is fluctuation in this type of variation to both directions, some of it not coinciding with coarticulatory phonetic phenomena. However, many enough examples of it are occurring in a phonetic environment that can be connected to coarticulation, which suggests that coarticulation is the reason behind the variation (see Dahlgren et al. 2022 for statistical co-occurrences). In addition, much of the variation between *eta* and other variants included in iotacism often display coarticulation. In (5), the standard *eta* has indeed been replaced with *iota*, an often-occurring feature of iotacism, but in the vicinity of /l/, which, again, is a coronal consonant and can thus cause a fronting effect on the vowel.

(5) ξυλωπωλις < ξυλοπώλης (O.Narm., 21; Dahlgren 2017: 104)

As can easily be imagined, the various misspellings that can be linked to iotacism are the most frequently occurring group regarding nonstandard vowel orthography in Greek in Egypt: variation between ι, υ, η, ει, οι <i, y, ē, ei, oi> (Dahlgren 2017). Of course, consonant-to-vowel coarticulation has been noted before by Coptologists (Kahle 1954; Girgis 1965). It may be a mistake to treat the apparent iotacism in Greek in Egypt as part of the Greek development as it seems so similar to the vowel variation in one of the related languages to Eyptian-Coptic, Arabic, and also shows the same type of contextual variation in documentary papyri from other areas besides Egypt, such as Palestine, again with regard to speakers from related languages (Dahlgren accepted). Furthermore, coronal consonants, those that can raise the vowel quality, are the biggest group of phonemes, and it has been considered that the Egyptian Greek vowel raising might have been caused due to the presence of coronals around them (Horrocks 2010: 168). Therefore, there was a conspiracy of sorts going on in Egyptian Greek, with pressure on especially the front vowels coming from both languages. According to Gignac (1991: 187), iotacism was accelerated in Egyptian Greek due to (Egyptian-)

---

[7] Note, however, that liquids (/l, r/) also frequently adapt to the quality of phonemes even one syllable away (Eriksson, Dahlgren and Vierros (2025: 9–14), but in this instance the liquid phoneme is the very next one coming up after the initial vowel phoneme.

Coptic having had fewer front vowels in the vowel inventory, which would have led to careless or uncertain qualities of especially those vowels that did not exist in (Egyptian-)Coptic (such as /y/). But this is probably just half the story, and the effect was further aided by the L1 tendency of Egyptian-Coptic for consonantal coarticulation, as well as the advancement of Greek-internal vowel raising in the Hellenistic era.

4.2 Stress-related phenomena

In another type of variation to the examples of underdifferentiation above, there are two stress-related transfer phenomena from Egyptian-Coptic to Greek: the replacement of /u/ with /o/ and vice versa, and the confusion of /a, e, o/ with each other, representing (in graphemic form) the reduction of these sounds in word-final position to schwa. Starting with /o, u/ variation, the standard ω /o/ has been replaced with ου /u/ in the nonstandard production of Μακρινου (6).

    (6) Μακρινου < Μακρίνῳ (O.Narm., 92; Dahlgren 2017: 84, 91)

This position is unstressed, both in Greek and would be according to Coptic prosodic rules, as well (Peust 1999) – therefore, this is basic phoneme redistribution related to stress, as described by Weinreich (1968). We know from the phonological descriptions of Coptic that /o, ɔ/ were not allowed in an unstressed position, while /u/ was (Peust 1999: 250–4; see also Dahlgren 2017: 83–90). There are instances to the other direction as well in several papyri, all matching Coptic phonological rules (Henriksson, Dahlgren and Vierros 2025: 13–7), so there remains little doubt that L1 stress rules were used in the L2 Greek production of some Egyptian speakers. Although the example shows variation in the word-final vowel quality, Gignac (1976: 211) noted that most of this type of variation occurred in word-initial or -medial position, which is exactly where the stress position would most often be affected in Coptic.

    By contrast, another stress-related phenomenon related to the variation between /a, e, o/ mostly occurred in the word-final position and can be seen in e.g. O.Narm. 115. The standard ο /o/ in the word-final syllable has been replaced with ε /e/ in a word regarding the mixing of wine, in what appears to be a private letter (7).

    (7) κερασεν < κέρασον (O.Narm., 115; Dahlgren 2017: 62 ff.)

In this case, the Egyptian-Coptic prosodic system seems to have been transferred onto the Greek word, apparently leaving the final syllable -ov vulnerable to neutralisation to schwa. This is a feature known of Coptic: the word-final syllable was often unstressed, and unstressed vowels in a word-final position were often neutralised to schwa in Coptic (Dahlgren 2017: 62–6, Peust 1999: 251–253).

All the features discussed in this paper are paralleled in nonstandard spellings of Greek loanwords in Coptic, which mark them as L2 usage, and reveal transfer of Egyptian-Coptic phonological properties onto Greek (Dahlgren 2017). This particular type of variation, the neutralisation of word-final vowels to schwa, is a frequent type coming from both Egyptian as well as non-Egyptian writers, all using L2 Greek in a language contact scenario (see e.g. the papyrus corpora P.Fay., O.Claud.; Leiwo 2010, 2017; Dahlgren & Leiwo 2020). The exact process of how this type of contact-induced variation came about is less clear than with the examples of underdifferentiation, but there is evidence that L2 learners can be 'deaf' to perceiving L2 stress, affected by the stress pattern of their L1 (Kijak 1977). According to Matras, there is evidence of accommodation to L2 stress patterns, even replacement of L1 patterns by them; however, Matras also states that adult L2 learners particularly have problems with leaving behind their L1 prosody, it being one of the most prevalent factors behind what is considered a foreign accent (Matras 2009: 69–70). What is definitely clear, however, is that this type of variation must be more speaker- than listener-induced as it may be possible to not hear what is there, as in the case of the foreign phoneme /y/, but it surely should be much more difficult to hear things that are not there. And in this case, the Egyptian L2 Greek users are replacing the Greek word-final vowel qualities /e, a, o/ with schwa, which is simply not reflecting the phonetic reality of the Greek original phonemes. Coptic had all of these vowel qualities (Peust 1999: 201), even if it has been claimed that no Coptic dialect had more than two of these at one time (Gignac 1991: 187). It is far more likely that whatever was heard, was replaced by the pronunciation of the L1 phonological content: the stress peak in Coptic apparently being strong (Horrocks 2010: 112, 169–70), the posttonic syllable would have been accompanied by a very weak vowel quality, expressed as schwa. This is the speaker replacing the target language's phonological features with those of their own, not unlike in e.g. Indian English, which uses full vowel qualities in all positions, and has replaced the English stress-timed stress system with a syllable-timed one, both features deriving from the Indian native languages (Dahlgren 2022: 139). For both of these language situations, it is possible that the language was in part learned from written form, seeing the

letters used for sounds and pronouncing them according to the phonological rules of one's L1. This is different from the underdifferentiation of Greek /y/ as /u/, for instance, as that must have been heard; Coptic alphabet does not include the letter *ypsilon*, because there was no sound corresponding to it, so the mispronunciation of Greek /y/ could not have been based on the letter-sound correspondence in the L1 writing system/phoneme inventory.

5 Summary and conclusions

As has been shown in this Chapter, the nonstandard variation sometimes present in the Greek papyrological documents can derive from a contact linguistic context and be the result of L1 (in this case, Egyptian-Coptic) transfer through more than one transfer phenomenon; to reduce the phenomena to only two, one can speak of the effect of underdifferentiation of foreign phonemes and stress-related transfer, including a change in the stress position and redistribution of phonemes related to the stress patterns. Probably the most surprising feature to match this description concerns iotacism but also the variation of the rounded vowels /o, u/ is relevant here. In L1 Greek, the latter mostly concerns case merger, whereas in the Egyptian L2 Greek material it can be related to case merger as well as solely the phonetic-phonological level, concerning vowel reduction in unstressed positions. However, in this chapter the focus was on how this type of variation came about in the first place: by the listener repeating things they heard in a compromised manner, affected by what was available in their L1, or by simply mispronouncing foreign phonemes the speaker had little experience with. It seems that in the underdifferentiation of phonemes, one could perhaps assign more weight on what was misheard, while in the stress-related variation, with its rather active replacement of positional allophones, one was changing the words' phonological forms so much that perhaps we can speak of speaker-induced variation, regardless of what was heard – in other words, integration of the foreign phonemes into the native language phonological system. To give a very simplistic example of a modern language contact situation, the Finnish L2 users of English do hear that the English loanword so frequently used in Finnish, *cool*, has aspiration in its original form, but it is usually not repeated in code-switching situations within Finnish conversations because Finnish does not have aspirated plosives. Therefore, the change of form in this loanword production is not accidental or unnoticed, but more actively and consciously produced to integrate it to the surrounding linguistic reality.

References

Greek and Coptic manuscript editions

OGN I = O.Narm. = Pintaudi, Rosario & Pieter J. Sijpesteijn. 1993. *Ostraca greci da Narmuthis (OGN I)*. Quaderni di Medinet Madi 2. Pisa: Giardini Editori e Stampatori in Pisa.

P. Hamb.bil. 1 = Diebner, Bernd J. & Rodolphe Kasser (eds.). 1989. *Hamburger Papyrus bil. 1. Die alttestamentlichen Texte des Papyrus Bilinguis 1 der Staats- und Universitätsbibliothek Hamburg: Canticum canticorum (coptice), Lamentationes Ieremiae (coptice), Ecclestiastes (graece et coptice)*. Genève: Cramer.

Bibliography

Allen, W. Sidney 1968. *Vox Graeca: A guide to the pronunciation of Classical Greek*. Cambridge: Cambridge University Press.

Bagnall, Roger S. 2007. 'Reflections on the Greek of the Narmouthis Ostraca', in Mario Capasso & Paola Davoli (eds.), *New archaeological and papyrological researches on the Fayyum, Proceedings of the International Meeting of Egyptology and Papyrology, Lecce, June 8th–10th 2005*. *Papyrologia Lupiensia* 14, pp. 14–21.

Bellem, Alex 2007. Towards a comparative typology of emphatics. University of London, doctoral dissertation.[8]

Catford, John Cunnison 2001. *A practical introduction to Phonetics*. Oxford: Oxford University Press.

Dahlgren, Sonja 2022. 'Egyptian Greek – A contact variety', in Bianconi, Michele, Marta Capano, Domenica Romagno & Francesco Rovai (eds.), *Ancient Indo-European languages between linguistics and philology. Contact, Variation, and Reconstruction*. Leiden: Brill. pp. 115–2.

Dahlgren, Sonja & Martti Leiwo 2020. 'Confusion of form or phoneme?', in Rafiyenko, Dariya & Ilja A. Serzant (eds.), *Postclassical Greek: Intersections of linguistics and philology*. Trends in Linguistics. Studies and Monographs. [TiLSM] Book 335. Berlin: De Gruyter. pp. 283–301.

Dahlgren, Sonja 2020. 'The system of Coptic vowel reduction: Evidence from L2 Greek usage', in Kuznetsova, Natalia & Cormac Anderson (eds.), *The*

---

[8] Single-spaced, re-paginated version (PDF) made by the author.

dynamics of vowel reduction and loss: Phonetic mechanisms and outcomes for phonology and morphology (Special issue on vowel reduction). *Italian Journal of Linguistics*, pp. 211–27.

Dahlgren, Sonja 2017. Outcome of long-term language contact: Transfer of Egyptian phonological features onto Greek in Graeco-Roman Egypt. University of Helsinki, doctoral dissertation.

Dahlgren, Sonja 2016. 'Towards a definition of an Egyptian Greek variety', *Papers in Historical Phonology* 1: 90–108.

Dahlgren, Sonja Accepted. 'Iotacism in Egyptian Greek: Greek-internal development and transfer effects from Egyptian-Coptic', in Bentein, Klaas & Marja Vierros (eds.), *New light from the East*. Helsinki: Helsinki University Press.

Dahlgren, Sonja, Alek Keersemaekers & Joanne Stolk 2022. 'Language contact in historical documents: the identification and co-occurrence of Egyptian transfer features in Greek documentary papyri', *Journal of Historical Sociolinguistics* 8(2): 325–56.

Fewster, Penelope 2002. 'Bilingualism in Roman Egypt', in J.N. Adams, Mark Janse, Simon Swain (eds.), *Bilingualism in ancient society: Language contact and the written text*. Oxford: Oxford University Press, pp. 220–45.

Gallo, Paolo 1997. *Ostraca demotici e ieratici dall'archivio bilingue di Narmouthis II*. Quaderni di Medinet Madi 3. Pisa: Giardini Editori e Stampatori in Pisa.

Gignac, Francis T. 1991. 'Phonology of the Greek of Egypt, influence of Coptic on the', in Atiya, Aziz S. (ed.), *The Coptic Encyclopedia*. Vol. 8 of 8. New York: Macmillan Publishers, pp. 71–96.

Gignac, Francis T. 1976. *A grammar of the Greek papyri of the Roman and Byzantine Periods*. Vol. I Phonology. Milano: La Goliardica.

Girgis, W.A. 1966. 'Greek loan words in Coptic (Part II)', *Bulletin de la Société d'Archéologie Copte* 18: 71–96.

Grossman, Eitan & Tonio Sebastian Richter 2017. 'Dialectal variation and language change: The case of Greek loan-verb integration strategies in Coptic', in Grossman, Eitan, Peter Dils, Tonio Sebastian Richter & Wolfgang Schenkel (eds.), *Greek influence on Egyptian-Coptic: Contact-induced change in an ancient African language*. DDGLC Working Papers I, *Lingua Aegyptia, Studia Monographica*, 17. Hamburg: Widmaier Verlag, pp. 207–236.

Halle, Morris 1962. 'Phonology in generative grammar', *Word* 18: 54–72.

Haugen, Einar 1950. 'The Analysis of linguistic borrowing', *Language* 26(2): 210–231.

Henriksson, Erik, Sonja Dahlgren & Marja Vierros 2025. 'Phonological variation in Greek papyri: Two case studies using PapyGreek search', in Reggiani, Nicola (ed.), *Digital Papyrology*. Vol. 3. Berlin: De Gruyter.

Horrocks, Geoffrey C. 2010. *Greek: A History of the language and its speakers*. Oxford: Wiley-Blackwell.

Jespersen, Otto 1949. *Efficiency in linguistic change*. Copenhagen: Ejnar Munksgaard.

Kahle, Paul E. 1954. *Bala'izah. Coptic texts from Deir El-Bala'izah in Upper Egypt*. Vol. I. London: Oxford University Press.

Leiwo, Martti 2020. 'L2 Greek in Roman Egypt: Intense language contact in Roman military forts', *Journal of Historical Sociolinguistics* 6(2): 1–31. https://doi.org/10.1515/jhsl-2019-0009.

Leiwo, Martti 2017. 'Confusion of mood and tense in Greek private letters of Roman Egypt', in Mark Janse & Klaas Bentein (eds.). *Linguistic variation and change: Tense, aspect and modality in Ancient Greek*. Amsterdam Studies in Classical Philology. Leiden: Brill, pp. 242–60.

Leiwo, Martti 2010. 'Imperatives and other directives in the letters from Mons Claudianus', in Trevor V. Evans & Dirk D. Obbink (eds.), *The language of the papyri*. Oxford: Oxford University Press, pp. 97–119.

Major, Roy C. 2001. *Foreign accent: The ontogeny and phylogeny of second language phonology*. Second Language Acquisition Research Series. Mahwah: Lawrence Erlbaum Associates.

Martinet, André 1964. *Elements of general linguistics*. Chicago: University of Chicago Press.

Matras, Yaron 2009. *Language contact*. Cambridge Textbooks in Linguistics. Cambridge: Cambridge University Press.

Ohala, John 1981. 'The listener as a source of sound change', in Carrie S. Masek, Roberta A. Hendrick, Mary Frances Miller (eds.), *Papers from the Parasession on Language and Behavior*. Chicago: Chicago Linguistic Society, pp. 178–203.

Peust, Carsten 1999. *Egyptian phonology. An introduction to the phonology of a dead language*. Göttingen: Peust und Gutschmidt.

Quaegebeur, Jan 1991. 'Pre-Old Coptic', in Atiya, Aziz S. (ed.), *The Coptic Encyclopedia*. Vol. 8 of 8. New York: Macmillan Publishers, pp. 190–1.

Reetz, Henning & Allard Jongman 2009. *Phonetics. Transcription, production, acoustics, and perception*. Malden: Wiley-Blackwell.

Sankoff, Gillian 2004. 'Linguistic outcomes of language contact', in Chambers, Jack K. & Natalie Schilling (eds.), *The handbook of language variation and change*. Hoboken, N.J.: Wiley, pp. 638–68.

Schleef, Erik & Ramsammy, Michael 2013. 'Labiodental fronting of /θ/ in London and Edinburgh: a cross-dialectal study', *English Language & Linguistics* 17(1): 25–54. doi:10.1017/S1360674312000317.

Strange, Winifred (ed.). 1995. *Speech perception and linguistic experience: Issues in cross-language research*. Baltimore: York Press.

Strange, Winifred & Valerie L. Shafer 2008. 'Speech perception in second language learners: The re-education of selective perception', in Edwards, Jette G. Hansen & Mary L. Zampini (eds.), *Phonology and second language acquisition*. Studies in Bilingualism. Amsterdam: John Benjamins Publishing Company, pp. 153–91.

Teodorsson, Sven-Tage 1977. *The phonology of Ptolemaic Koine*. Göteborg: Acta Universitatis Gothoburgensis.

Thomason, Sarah G. 2001. *Language contact. An introduction*. Edinburgh: Edinburgh University Press.

Thomason, Sarah Grey & Terrence Kaufman 1992. *Language contact, creolization, and genetic linguistics*. Berkeley: University of California Press.

Tollfree, Laura 1999. 'South East London English: discrete versus continuous modelling of consonantal reduction', in Folkes, Paul & Gerard Docherty (eds.), *Urban voices*. London: Arnold, pp. 163–84.

Torallas Tovar, Sofia 2010. 'Greek in Egypt', in Bakker, Egbert J. (ed.), *A companion to the Ancient Greek language*. Hoboken, N.J.: Blackwell, pp. 253–66.

Traunmüller, Hartmut 1999. 'Coarticulatory effects of consonants on vowels and their reflection in perception', in *Proceedings from the XIIth Swedish Phonetics Conference*, pp. 141–4.

Vieru-Dimulescu, Bianca & Philippe B. de Mareüil 2006. 'Perceptual identification and phonetic analysis of 6 foreign accents in French', in *Interspeech 2006 – ICSLP, Ninth International Conference on Spoken Language Processing, Pittsburgh, PA, USA, September 17–21, 2006*, pp. 441–4. ISCA Archive,http://www.isca-speech.org/archive/interspeech_2006.

Weinreich, Uriel 1968 [1953]. *Languages in contact. Findings and problems.* The Hague: Mouton Publishers.

Zimmer, Karl & Orhan Orgun 1992. 'Turkish'. *Journal of the International Phonetic Association* 22(1–2): 43–5. doi:10.1017/S0025100300004588.

2. Scribal Revision in the Process of Text Production. A Linguistic Typology of Scribal Corrections in Four Genres of Greek Documentary Papyri[1]

Joanne Vera Stolk

1 Introduction

Scribal revision gives us an opportunity to observe the scribe at work and obtain closer insights into the role of the scribe in the process of text production. Scribal corrections are usually noted in Greek papyrus editions by applying brackets in the text or comments in the *apparatus criticus*, but the phenomenon has not been studied comprehensively. Papyrologists often regard the presence of corrections as an reason to identify the text as a draft, as, for example, Sijpesteijn and Worp (1977: 91), who conclude about a papyrus from the Vienna collection: 'The many deletions and interlinear additions indicate that we are dealing with a rough draft.' A draft, in this sense, means 'a preliminary sketch or rough form of a writing or document, from which the final or fair copy is made'.[2] While the presence of scribal revision might seem a good indication of drafting, this principle may not apply to all genres of documentary papyri in the same way, as Luiselli (2010: 73–4) remarks:

> 'Evidence of extensive textual reworking is usually treated as an indicator of a draft, whether the text is a literary composition, a contract, a private letter, or a petition. But fair copies of letters are more likely than the vast majority of petitions to display a reasonable number of corrections, so that it may not be easy to distinguish a draft of a letter from a fair copy.'

[2] See 'draft, n.' s.v. 5 in the *Oxford English Dictionary Online*, Oxford University Press, March 2022, www.oed.com/view/Entry/57398. Accessed 4 May 2022.

Luiselli suggests here that final versions of letters are more likely to have corrections than final versions of petitions and that corrections can thus not straightforwardly be interpreted as an indicator of a draft in letters. This presence of corrections in final versions of letters is visible, for example, in the letters of the Zenon archive (TM ArchID 256). Many of them contain corrections, even though most of these papyri are final copies that have been sent to and received by the protagonist Zenon. Some of these corrections seem to have been produced during writing, as in πλήθει (ε *corr. ex* ι) in P.Col. 3, 8, 5, where the ε was written over the previous ι and the final ι added, while others may have been inserted even after the text was finished, such as the repeated insertion of the ε above the line in the phrase ἐμ πόλ\ε/ι in P.Cair.Zen. 3, 59301, 2 and 5 (see Stolk 2019).[3] Corrections may thus be introduced in final versions of letters, but does this mean that corrections are also more commonly found in letters compared to other genres? Papathomas (2018) has shown that the corrections in papyrus letters may apply to different levels of language organization, like spelling, grammar or syntax. Can these different types of corrections be found in all genres in equal measures? Or could the linguistic level of the corrections perhaps also tell us something about the method and stage of composition of a document?

As scribal corrections in papyri have not been studied on a large scale before, I will first give an chronological overview of the presence of scribal corrections in several genres of documentary papyri (Section 2). Next, I will categorise the examples of scribal revision linguistically according to the linguistic unit the correction applies to and show the distribution of these different types of corrections across the genres (Section 3). These quantitative results are complemented by a qualitative study of corrections in several papyrus archives dating to the Roman and Byzantine periods in Egypt (Section 4), such as the archives of an Alexandrian scribal office (late first century BCE), the police chief of Euhemeria (first century CE), the scribal office in Tebtynis (first century CE),

---

[3] Papyrus editions are cited according to the Checklist of Editions of Greek, Latin, Demotic and Coptic Papyri, Ostraca and Tablets, at www.papyri.info/docs/checklist. The Greek text and metadata are based on the digital edition available in the Papyrological Navigator (www.papyri. info) and checked against the *editio princeps*. The use of critical signs is in accordance with the so-called 'Leidener Klammersystem' (cf. Van Groningen 1932: 262–9). Scribal deletions are indicated by double square brackets ⟦ ⟧, scribal insertions by slashes \ /. Text between single square brackets [ ] is not preserved on the papyrus, but supplemented by the modern editor; dots under letters signal uncertain readings by the editor. Notes from the critical apparatus are here inserted between brackets in the Greek text ('*corr. ex.*' provides the form from which the text is corrected on the papyrus and '*l.*' signals a regularization by the modern editor). Translations are my own, but they may be based on the translation of the edition, if available.

the governor Apollonios (second century CE), the village scribe Petaus (second century CE), Aurelius Ammon (fourth century CE), Dioscorus of Aphrodito (sixth century CE) and the Apion family (fifth to seventh centuries CE). Finally, I will reflect on the information that scribal corrections could provide about the method and stage of text production (Section 5).

2 Scribal corrections in different genres

The corpus for this study consists of all published documentary papyri with a digital edition in the Duke Databank of Documentary Papyri (DDbDP), which is accessible through the Papyrological Navigator (www.papyri.info). The Greek texts (state January 2014) were imported by Mark Depauw and editorial regularizations isolated from the texts (see Depauw and Stolk 2015).[4] Similarly to the editorial regularizations, the corrections by ancient scribes are usually marked in the edition. Editors apply double square brackets (⟦α⟧) to indicate deletions, slashes (\α/) for text written above or below the lines and comments of the type 'α *corr. ex* β' in the *apparatus criticus* to indicate changes to the text made in antiquity. These types of scribal corrections have also been retrieved from the digital editions by Mark Depauw and have been annotated by the author of the present article within Trismegistos.

For this paper, corrections with an uncertain reading of the correction, the corrected form or the direct linguistic context as well as possible abbreviations of words (sometimes also indicated as insertions above the line) were removed from the corpus, resulting in a total of 20,717 corrections. In order to compare different types of documents, the genre or text type of every document was identified as belonging to one of the following groups: letters, contracts, declarations (including petitions), pronouncements, reports, receipts and lists. The general categorization was based on the information available in Trismegistos from previous studies, the subjects attached to each text in the *Heidelberger Gesamtverzeichnis der griechischen Papyrusurkunden* Ägyptens (*HGV*), additional information provided in *Advanced Papyrological Information System* (*APIS*), and the title of the original edition.[5]

---

[4] The results of this are available at www.trismegistos.org/textirregularities.

[5] A more detailed account of the categorization into text types (and subtypes) can be found in Stolk (2020). The resulting categorization is also accessible online through TM texts (www.trismegistos.org/tm/), see 'type'.

Excluding semi-literary texts and fragmentary documents of which the genre could not be determined, 7,993 of the remaining 46,376 documentary papyri (17%) seem to contain at least one correction. Four main genres have been selected for comparison in the following sections: (1) contracts, (2) letters, (3) lists and accounts and (4) petitions. The first category includes all types of contracts and juridical agreements; the second category includes all types of letters used for official, business and private correspondence. The third category, lists and accounts, is limited to itemised collections of information from both private and official contexts, thus excluding abstracts of contracts or registers of official correspondence that rather take the form of a collection of shorter and longer texts. The fourth category, petitions, includes various types of requests and complaints directed to persons in a higher position, but excludes general notifications addressed to the authorities, such as census applications or notifications of birth and death. Figure 1 provides a chronological overview of the presence of scribal corrections for each genre.[6]

Figure 1. Percentage of texts with corrections in four different genres of papyrus documents from the third century BCE until the seventh century CE.

---

[6] The chronological results in Figures 1 and 3 are weighted graphs in which papyri dated to more than one century are spread out over the time range they are dated to (cf. Van Beek and Depauw 2013). All results are based on published Greek papyri in the DDbDP (state January 2014) and the annotated database TM Text Irregularities (state March 2018).

The chronological distribution is partly dependent on the presence of archives preserving a group of documents produced under similar circumstances. For example, in the Zenon archive (TM ArchID 256), the archive of Menches and the village scribes of Kerkeosiris (TM ArchID 140) and the archive of the *Katochoi* of the Sarapieion (TM ArchID 119), the percentage of corrected lists is higher than in the remaining contemporary papyri. These results have an influence on the percentage of corrected lists in the third and second centuries BCE. Similarly, the lists in the archives of Apion (TM ArchID 15) and Dioscorus (TM ArchID 72) add slightly to the higher percentages for the sixth century. Petitions seems particularly vulnerable for the deviations posed by individual archives. For example, the higher percentage of scribal corrections for petitions in the second century BCE is mainly due to the archive of the *Katochoi* of the Sarapieion, while the peak in the sixth century CE is largely the result of frequent corrections in petitions of the Dioscorus archive. Furthermore, the archive called 'Petitions from Euhemeria' (TM ArchID 187) contributes to the peak in the first century CE and the petitions of the archive of Aurelius Ammon (TM ArchID 31) to the fourth century CE (see 4.2).

Leaving the chronological variation aside, some general differences between genres can be observed. On average, a lower percentage of corrected texts is found among the lists (19%) and contracts (22%), while corrections seem slightly more common in letters (28%) and petitions (32%).[7] Figure 1, however, does not indicate the number of corrections per text. The identification of a text as a draft is often based on the evidence of more extensive revision rather than the presence of a single correction. Figure 2 shows the number of corrections per text for each of the four genres.[8]

---

[7] I used Pearson's chi-squared test to determine whether there is indeed a significant difference between the results observed here and results that would have been generated by chance. Even though the differences between the genres are not enormous, the very low p-value shows that it is unlikely that the differences are caused by chance (chi-square = 281.53, 3 degrees of freedom, p < 0.00001). The standardised residuals of the chi-squared test show that the genres list, letter and petition contribute most to the chi-square value. The effect size of the results is small (Cramer's V = 0.11), which means that the factor 'presence of corrections' is probably not the best denominator of the differences between the genres. Overall, the differences are significant enough to suggest that these genres have some individual properties that would increase or decrease the likelihood of corrections appearing in the documents preserved to us.

[8] The given estimates are likely to be lower than the real numbers of corrections, since corrections with uncertain readings are left out in this study.

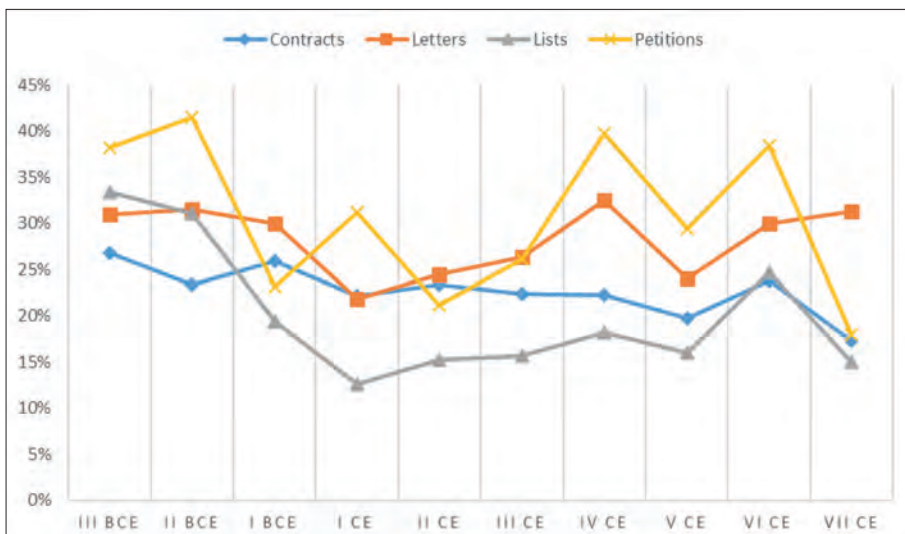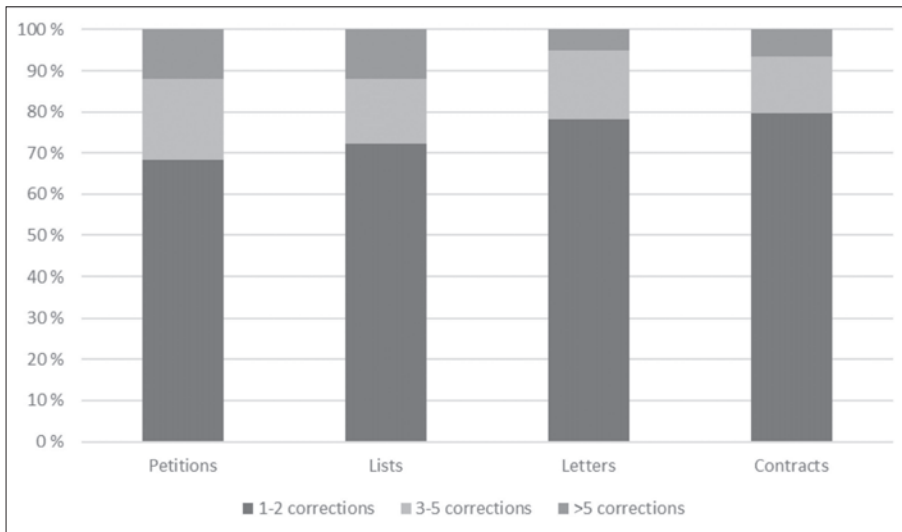Figure 2. Percentage of texts with 1–2, 3–5 or more than 5 corrections per text in four different genres of papyrus documents from the third century BCE until the seventh century CE.

Figure 2 shows that corrections in petitions tend to come in higher numbers than in other genres. While almost 80% of the corrected letters and contracts contain only one or two corrections and even 95% contain no more than five, corrections in petitions and lists tend to be more numerous with more than 10% containing more than five and around 30% containing more than two.[9] If the presence of corrections, especially in higher numbers per text, can be taken as a indicator of a draft, it is expected to find more drafts among petitions. These general differences between the genres will be examined in more detail in the following Section by distinguishing between different types of corrections.

3 Linguistic categorization of scribal corrections

Scribal revision involves a wide range of scribal activities: from extensive alterations to a document at an early stage in the process of composition to minor improvements to a finished text. The stage in the composition process is thus

---

[9] The differences between the genres are significant (chi-square 87.59, 6 degrees of freedom, p < 0.00001), although the effect size is small (Cramer's V = 0.09). The standardised residuals show that the numbers of petitions, lists and letters with more than five corrections contribute most to the chi-square value.

also expected to have an impact on the type of scribal corrections. For example, the corrections by Dioscorus (TM ArchID 72) to the petitions P.Cair.Masp. 1, 67002 and P.Lond. 5, 1674 take the shape of superlinear insertions of words and short phrases, while his emendations to the petitions P.Cair.Masp. 1, 67006 and 67020 concern mostly orthographic issues (Stolk 2019). In the first case, the documents are preliminary drafts produced by Dioscorus himself during the process of composition, whereas the other two documents are complete texts that have been reproduced by someone else with minor corrections added later by Dioscorus. Phrasal revision seems characteristic of the preliminary stages of free composition in these documents, while orthographic corrections are added at a later stage and/or following a different production method. Hence, linguistic categorization of scribal corrections may be helpful to identify different methods and stages of production of documentary papyri. Various motivations (e.g. stylistic, rhetorical, practical) for corrections in papyrus letters from the fifth to the eighth century CE have been identified by Papathomas (2018), but in order to compare a large number of corrections in various text types, we first need to define the general levels of linguistic analysis to which every correction could be assigned, before looking into more detailed motivations for corrections at those levels. All corrections in Section 2 have been categorised by the author of the present article according to the linguistic unit each correction applies to. I have distinguished the following four basic linguistic levels:

1) The **grapheme** or **phoneme level** contains deletions, insertions and changes to a grapheme (smallest unit of writing, i.e. one letter) or digraph (two letters) corresponding to one phoneme (unit of sound) or a diphone (two phonemes expressed by one character, such as ψ and ξ) in Greek, including corrections of gemination, simplification and metathesis (for these phenomena, see examples in Gignac 1976: 154–65; 314–15). There can be more than one correction of a grapheme or phoneme per word, but only when these are not forming one unit of morphological or lexical meaning (see below).

2) The **morpheme level** includes deletions, insertions and changes to a morpheme (unit of grammatical meaning), such as a case or verb ending. Morphemes consisting of one phoneme have been annotated for both grapheme and morpheme levels, but are counted here only as morphemes in order to avoid making an ambiguous decision in each case.

3) The **lexeme level** applies to deletions, insertions and changes to a full lexeme (unit of lexical meaning) or part of a lexeme that cannot be explained at a phonological or morphological level (see above). These changes may be meaningful, although

the circumstances do not always allow complete understanding of the semantic or syntactic change involved. Corrections effecting numerals and symbols are annotated as subcategories to the lexical level.

4) The **phrasal level** contains all deletions, insertions and changes of two or more words.
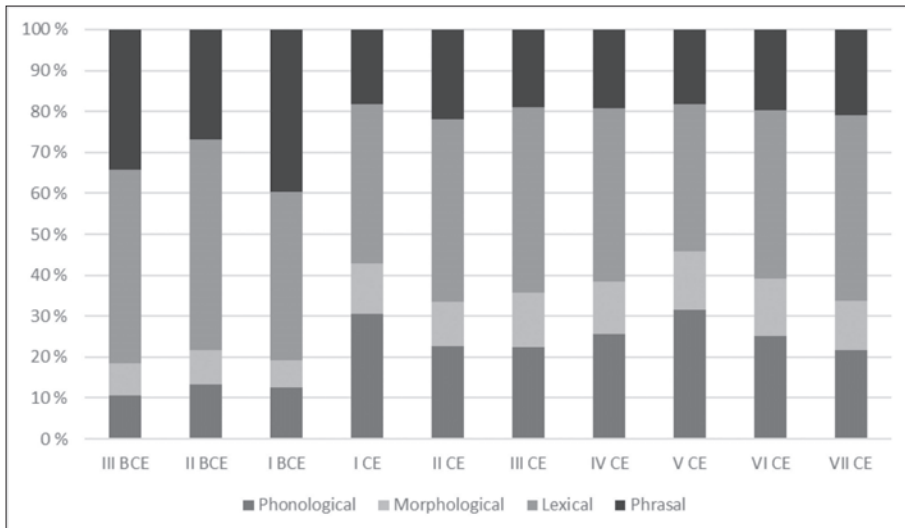

Figure 3. Proportional distribution of corrections according to their level of linguistic analysis from the third century BCE until the seventh century CE.

Figure 3 shows the distribution of the corrections according to the linguistic level for every century. A chronological difference needs to be pointed out first. While the first to seventh centuries CE show a similar pattern of around 20–30% corrections at the grapheme level, 10–15% at the morpheme level, 35–45% lexical and around 20% phrasal, the Ptolemaic period stands out with a generally lower level of grapheme and morpheme corrections, only 20% counted altogether, and a much higher proportion of phrasal revisions, around 30–40%.[10] It is difficult

---

[10] The differences between the Ptolemaic and Roman to Byzantine periods are significant (chi-square = 728.40, 3 degrees of freedom p < 0.00001), although the effect size is relatively small (Cramer's V = 0.19). The standardised residuals show that the numbers of grapheme and phrase corrections are contributing most to the chi-square value. This difference cannot be explained by a difference in the genres preserved from these periods: Ptolemaic contracts, letters, lists and petitions all contain a smaller portion of grapheme and morpheme corrections than the same genres in the Roman and Byzantine periods (apart from the percentage of morphological corrections in lists which is equally low for all periods). The chronological difference is most evident in contracts and

to point out a single factor explaining this change and any attempt to identify particular historical differences between these two periods will end up being speculative, lacking concrete (quantitative) evidence for it. One factor could be the phonological changes that start in the Ptolemaic period, but become much more extensive in the Roman and Byzantine periods, making the written language more difficult to spell. Spelling variation in general tends to be less frequent in documents from the Ptolemaic period in comparison to the Roman and Byzantine periods (Stolk 2020) and this may reflect on the felt need for corrections. It is likely, however, that there are several other factors involved as well, such as changes in the levels of education and literacy of the scribes involved or different attitudes towards spelling variation and corrections (Stolk 2019; see also Bucking 2007). In order not to let this chronological difference interfere too much with the other factors, the comparison of the linguistic levels of corrections across the genres in Figure 4 is only applied to documents from the Roman and Byzantine periods.
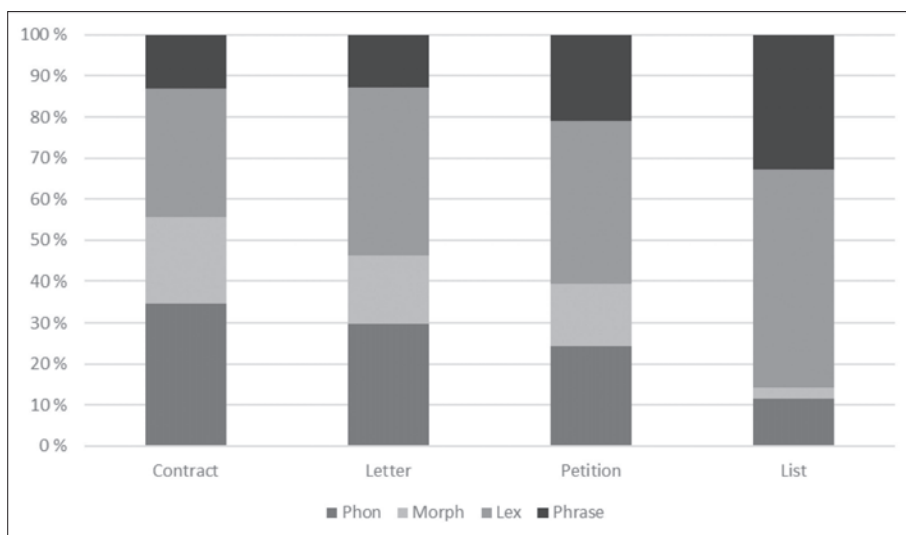
Figure 4. Proportional distribution of corrections according to their level of linguistic analysis within four different genres from the first until the seventh centuries CE.

There are clear differences between the genres with respect to the linguistic levels of the corrections. While the majority of the corrections in contracts affect

---

letters, where the percentage of grapheme and morpheme corrections in the Roman and Byzantine periods can be double or triple the amount in the Ptolemaic period.

graphemes or morphemes and only 13% affect phrases, 86% of the corrections in lists are phrasal or lexical and only 15% at the grapheme or morpheme levels.[11] Again, there are many possible reasons for this distribution. Linguistically, the relatively high percentage of grapheme and morpheme corrections in contracts and letters could be related to the complexity of inflection and (morpho)syntax in these genres, as opposed to lists which often involve references to single items lacking syntactic context and/or abbreviations omitting morphological information.

The absence or presence of corrections may also have been the result of the process and stage of composition of the documents that we have or the context of use. In order to elicit corrections, a particular difficulty to produce a form needs to coincide with a motivation to make emend. Very little is known about the acceptance or avoidance of mistakes or corrections in different genres. One could think that in documents meant for internal usage the presence of irregular forms was considered less important and corrections therefore less relevant. On the other hand, the presence of corrections itself could have been considered objectionable in more official documents, while the same corrections could have been regarded as acceptable or even desirable in more informal contexts. Other possible reasons for the differences between the number and type of corrections in these genres will be examined in more detail in individual texts from various archives in Section 4.

4 Scribal corrections in archives

The frequency of occurrence of nonstandard spellings in papyrus documents can differ according to the method and stage of composition. For example, the archive of the *Katochoi* of the Sarapieion (TM ArchID 119) contains various petitions and letters written in the hand of Apollonios, the younger brother of Ptolemaios (cf. UPZ 1). His petitions contain on average more nonstandard spellings than his letters, because the petitions are preliminary drafts while the letters are his copies of official letters written by others or final versions of his own private letters (Stolk 2020). These different methods of production (copying or drafting) and stages of composition (preliminary or final) are likely to have an impact on the presence, number and type of corrections as well. Knowledge

---

[11] The differences between the genres are significant (chi-square = 1310.64, with 9 degrees of freedom, p < 0.00001), although the effect size is relatively low (Cramer's V = 0.22). The standardised residuals show that the differences between contracts and lists contribute most to the high chi-square value.

about the production circumstances is more easily available and comparable in a group of related documents, such as a papyrus archive. In this Section, I will describe the background of writers and their methods of text production in more detail in order to study more closely the presence and distribution of different types of corrections in several private and official archives from the Roman and Byzantine periods.

4.1 Corrections in final versions of letters

Letters tend to contain corrections at all linguistic levels (see Figure 4), although in the far majority of the cases, we find only a limited number of them in one text (see Figure 2). The archive of the governor Apollonios (TM ArchID 19) contains more than two hundred administrative and personal documents collected during his time as the governor (*strategos*) of the district of Apollonopolites Heptakomias in Egypt (113–19 CE). The majority of these documents are letters, including a large number of private letters sent to him by his family in Hermopolis. Most of the letters, therefore, can be considered final versions received by him rather than personal drafts of outgoing documents. Still, about a third of these letters contain corrections, albeit in modest quantities: half of them have only one and none has more than five.

The corrections are found in letters by writers with various backgrounds: from beginners (cf. the alphabetic hand in P.Brem. 22) to more experienced scribes (cf. the chancellery style in P.Brem. 5). The methods of revision also vary. When the correction concerns only one letter, the old letter could be adapted, such as the η changed into ει in εἰδώς (P.Giss. 1, 45, 7), or the new letter just written over the old one, as the υ written over the second ζ in ἐπιζεζξῃς (P.Brem. 5, 12) and the first ρ in πρπρασκεται changed into ι by applying a thick vertical stroke (P.Brem. 22, 9). The most common methods are deletion by stroke(s), insertions of words and letters above the line or a combination of both to indicate a replacement. A more sophisticated method is found in two letters sent by Epaphrodeitos, where the writer implicitly deletes a letter and word by placing other letter(s) above it, cf. the replacement of γράψειν ('to write') by γράψει\ς/ ('you will write') and μοι ('me') by \αὐτῇ/ ('her') (P.Giss.Apoll. 22, 6–7 and 9). In the same way, he deletes letters by putting short diagonal strokes above each letter instead of any new letters, see e.g. the strokes above γνώμης σου ('your judgement') (P.Giss.Apoll. 22, 20) and τό (P.Giss.Apoll. 23, 4).

The examples of revision in the letters addressed to Apollonios concern mostly minor changes to graphemes, morphemes and short words. The changes

to graphemes are not caused by phonological merger only. For example, the first ρ in πρπράσκεται corrected to πιπράσκεται ('it has been sold') (P.Brem. 22, 9) could rather be explained by omission of the reduplication in anticipation of the next syllable. The spelling of <αι> instead of <ε> in διεβ⟦αι⟧\ε/βαι|ωσαμ[ην] ('I guaranteed') (P.Brem. 5, 10–11) does probably entail a phonologically motivated interchange (identical pronunciation of <ε> and <αι>), but in this particular context, it may also be enhanced by anticipation of the spelling <αι> in the following syllable. Regarding the morphemes, there are examples of paradigmatic merger as part of morphological changes, such as confusion between the first and second aorist endings (for the reasons behind morphological simplification of verb endings see Leiwo 2017) in προσῆλθα corrected to προσῆλθον ('I came') (P.Brem. 54, 4) and between the accusative singular of the i-stems and consonant stems (see more examples in Gignac 1981: 55–8) in Εὐδαίμονιν corrected to Εὐδαιμονίδα (P.Brem. 61, 21). A change of morphemes may also be motivated by a desire for reformulation. For example, the original ending of πρὸ πάντ⟦ων⟧ corrected to πρὸ παντ\ὸς/ ('before all') (P.Brem. 61, 16) is morphologically perfectly fine, but the sender or writer wanted to introduce an alternative that (s)he deemed more suitable in this context. Reformulation is also an important motivation for changes to larger elements. The deletion, insertion and replacement of (part of) words and short phrases could be reactions to scribal errors, such as skipping \λάβῃ/ in ὅπως παρα\λάβῃ/ παρ' ἐμοῦ ('so that he takes over from me') (SB 10, 10278, 15), but in other contexts these larger revisions may indicate attempts to rethink or rephrase the content of the letter.

Changes, even those in formulation and content, can be made during and after the process of composition of the (final) letter. Diskas started a greeting formula with ⟦ἀσπάζομαί σε⟧ ('I greet you') (P.Brem. 16, 52), but then realised that he first wanted to say something else and removed the greeting. Also Kornelios changed his mind about what he wanted to tell Apollonios (P.Giss. 1, 65, 9–10). He started a new sentence with ἀντέστη [δ'] ἐμοὶ ὁ τῆς [κώμης] | πράκτωρ ⟦φάσκων ο⟧ 'the tax collector of the village(?) was set against me saying …', but then he removed the introduction of the quote ('saying') and started a new sentence. Other corrections may have been made by the writer upon rereading previous sentence(s) or even after the whole letter was finished. The writer of P.Giss.Apoll. 37 thought that he had forgotten the infinitive ἔχειν ('to have') and added it above the line in l. 5, only then to realise that the infinitive was already present at the end of l. 4 and to remove the insertion again. The same letter could preserve evidence of changes made during and after writing. The writer of SB 10, 10278, 15, deleted a superfluous σε ('you') in l. 2 and inserted

\λάβῃ/ in l. 15 (see above) probably after writing these sentences, but he also seems to have overwritten a mistaken start of a following word beginning with μ (probably μου) by the final ε of the vocative ending in κύριε ('lord') (l. 6) and to have changed αὐτ⟦ά⟧ ('these things') into αὐτῷ ('him') by inserting an ω after the deleted α (l. 16). These last two corrections are more likely to have taken place during writing. In the same way, the ε is written above the αι in διεβ⟦αι⟧\ε/βαι|ωσαμ[ην], while the υ is written over the ζ before continuing with the ξ in ἐπιζεζξης (P.Brem. 5, 10–11 and 12). By far the majority of these corrections seem to have been undertaken by the same scribe who wrote the letter in the first place (see also Papathomas 2018: 163–6), although the initiative for the changes, especially those added later, could have been taken by someone else, such as the client or a supervisor. Occasionally, a second hand, possibly of the sender/author of the letter, is responsible for some final changes, such as perhaps the insertion of the enforcing adverb ἀεί ('always') in P.Giss.Apoll. 21, 10.

These minor corrections to final versions of letters testify of a fluid composition process. As (private) letters can be composed freely and preferably without wasting papyrus on numerous drafts, mistakes are easily made and stay visible to the addressee. People also tend to change their mind about the precise formulation or even contents of the message they want to convey during the process of composition. Spontaneous addition of extra lines in the margins of private letters attest of a similar phenomenon (see Homann 2012). Corrections added later show that many writers or authors may have reread their letters during or after writing to check for mistakes in orthography, morphology and syntax. Clearly, they cared about the language of the final product and a limited number of corrections was to be preferred above giving a wrong impression or leaving unintended linguistic irregularities.

4.2 Production of contracts

Contracts contain relatively few corrections (Fig. 2) and most of them seem to affect graphemes and morphemes (Fig. 4). This can also be observed from the corrections in contracts in the archive of the Apion family, dating to the sixth and seventh centuries CE (TM ArchID 15). The contracts are signed by the notary and have endorsements on the *verso*, so that we may safely assume they contain the final version of the document. The majority of the corrections affect graphemes and morphemes and there are no corrections at a phrasal level. For example, in P.Oxy. 16, 1970, 30, the last letter of Ἀνοῦπ was first written as α, perhaps in anticipation of the following patronymic Ἀνδρέου, with π written

over afterwards. The correction is found in the subscription written in a less formal style than the body of the contract. Another minor correction is found in the subscription by the agreeing party written in his own hand in P.Oxy. 24, 2420, 21. The names Παπνουθίου καὶ Ἀροθίου (genitive) seem to have been written in accusative (-ον) at first, perhaps modelled on the form in which they occur in the body of the document (l. 11). Corrections are also found in the more formal body of the contract. For example in P.Oxy. 1, 138, 28, the scribe started writing παν, before realizing the gender of the following noun and correcting it (probably immediately) into πᾶσαν χρείαν ('all needs').

As the contracts in the Apion archive illustrate, papyrus contracts are often final documents that have been kept by the parties in their (personal) archives. Changes to the formulation and contents are rare in these final copies and they may even have been unacceptable, since they could interfere with the legal validity of the product. Changes to phonemes and morphemes, however, do not seem to pose a major problem and are regularly found. Juridical phrases could be produced with the help of model formularies (Bucking 2007). The semantic and syntactic complexity of these precomposed phrases could have caused difficulties for scribes who may not have been able to compose a document like that without the help of models (see also the variation in the contracts from Pathyris in Vierros 2012a and 2012b). Uncertainty about the choice and spelling of morphological endings (see Leiwo 2017) seems to be the reason for most of the corrections in the Apion archive. In P.Oxy. 1, 135, the ω's in τω αυτω κτῆμα ('the same building'), εἰς ἕτερων τόπων ('to another place') and ἐπιζητούμενων | αὐτων ('him being required') (ll. 20–22) had to be changed into ο's to form the expected accusative single case endings. Since all of these corrections concern the same feature, it is likely that the mistake was only discovered after the text had been finished, possibly even by someone else. Nonstandard orthography is very common in contracts and most of the time variation seems to have passed unnoticed by the scribes (Bucking 2007; Stolk 2020). While standard spelling may not have had the highest priority in contracts, the numerous examples of orthographic and morphological corrections show that it was considered relevant, at least to some scribes and notaries, and that these types of corrections could be added without compromising the validity of the final product.

Although the quantitative results and the previous examples may give the impression that all contracts were produced based on fixed models without variation in textual composition and only minor variations in orthography and morphology, this was probably not the case. Relatively few drafts of contracts survive, but they do exist. The archive of the Alexandrian scribal office of a

legal specialist (TM ArchID 430) provides some examples of revised versions of contracts from the early years of the Roman period (Seidl 1973: 67 no. 2.1).[12] Most of the corrections are found at lexical (38%) and phrasal (49%) levels, such as the interchange of the order of the months Hathyr and Pachon in ἐν μὲν τῶι ⟦Παχ[ὼν]⟧ \Ἀθὺρ/ [(δραχμὰς) σ, ἐν] δὲ τῶι ⟦Ἀθὺρ⟧ \Παχὼ(ν)/ ἄλλας (δραχμὰς) σ ('in ~~Pachon~~Hathyr 200 drachmas, and in ~~Hathyr~~Pachon another 200 drachmas') (BGU 4, 1132, 34). Word order is a common topic for revision, see for example the change of the noun from pre- to postadjectival position in ἐκ τοῦ ⟦κλήρου⟧ Ἱεροξένο(υ) | κλήρου ('from the allotment of Hieroxenos') (BGU 4, 1167, 73–4). These revisions could be the result of copy mistakes (anticipation) or rethinking the formulation during writing and they are comparable to the continuous corrections and improvements found in drafts of petitions (see Section 4.3 below).[13] The process of textual composition can be followed more closely in the case of hesitations by the scribe, such as the later insertion and subsequent deletion of the names \⟦παρὰ Σα(ραπίωνος)⟧/ and \⟦Μάρκου καὶ Ἰσιδώρας⟧/ (BGU 4, 1149, 13–14). Some changes seem to have been made immediately during writing, such as the anticipation of ἐκ, which is removed to insert τῶι Γαίωι, in τῆς πρ(άξεως) γεινο(μένης) ⟦ἐκ⟧ | τῶι Γαίωι ἔκ τε ἀμφ(οτέρων) ('the right of execution being with Gaius on both …') (BGU 4, 1122, 27–8). Others seem to have been made at least after the phrase was written down, such as the fronting of ἐκ τοῦ ἰδίου in \ἐκ τοῦ ἰδίου/ ⟦ταυτὰ γένη⟧ αὐτενίαυ(τα) ⟦ἐκ τοῦ ἰδίου⟧ ('these crops for one year from your own' changed to 'from your own for one year') in a phrase added above the line (BGU 4, 1122, 23a).

According to the editor of BGU 4, 1160, the correction of Τειμοκράτης from the short form Τειμᾶς (l. 2) suggests that the scribe had the parties telling their names in front of him while he was drafting. Although it seems doubtful to conclude this practice from the correction of a name only, a similar procedure can be reconstructed from the (parts of) contracts in the archive of Kronion, the head of the scribal office of Tebtynis during the middle of the first century CE (TM ArchID 93). A first version of the contract seems to have been drawn up at the scribal office and signed by the contracting parties. The subscriptions

---

[12] Unfortunately, there are no images available for the majority of the documents from this archive. My observations are based on the corrections mentioned in the editions in BGU 4.

[13] Fixed juridical formulas play an important role in the composition of contracts, but the revisions show that also juridical phraseology can be employed with minor variations, see e.g. the corrections around the *praxis* clause in BGU 4, 1175, 10–15. Confusion between variant formulations of juridical formulas may also lead to linguistic inconsistencies and corrections (see Vierros 2012b; Stolk 2015: 268–77).

by the parties were copied onto an empty sheet of papyrus, to which the final copy of the contract was added later and collected by the parties (Husselman, P.Mich. 5, pp. 3–11). P.Mich. 5, 340 contains the drafts of two contracts. The space for the physical characteristics of the parties has been left blank and the subscriptions by the parties are written by the same scribe as the body of the document. Changes in composition are made while drafting, such as deletion by encircling ⟦ἐν προσφορᾷ κατὰ τὴν⟧ | ⟦[συγγρα]φὴν⟧ ('as a gift in accordance with the contract') immediately followed by rephrasing as κατὰ τήνδε τὴν ὁμολογείαν ('in accordance with this agreement'), as well as later insertions above the line \ὑπὲρ τῆς προγεγραμμένη(ς) Ἡρακλείας/ ('for the benefit of the aforesaid Herakleia') (P.Mich. 5, 340, 8–9 and 10). Just like other more complex documents, the contents of a contract needed to be discussed with the clients and the text composed by a scribe or notary before the final version could be produced and copied.

This process of composition is characterised by the presence of corrections at lexical and phrasal levels in the drafts of contracts in the archives of the scribal offices in Alexandria and Tebtynis, in contrast to the minor corrections to graphemes and morphemes in the final copies of contracts in the Apion archive. The archive of the notary Dioscorus preserves petitions at preliminary and later stages of composition, corresponding to corrections at different linguistic levels (see Section 3). The same phenomenon is found in his contracts. P.Cair.Masp. 2, 67151, containing a version of the testament of Flavius Phoibammon, has been copied by a scribe from a draft written by Dioscorus himself (P.Cair.Masp. 2, 67152). As a faithful copy (see P.Cair.Masp. 2, pp. 101–2), there is no need for any lexical or phrasal revisions at this stage. All corrections affect changes to graphemes, morphemes and occasional insertions of small words. On the other hand, on the *verso* of a marriage contract (P.Cair.Masp. 3, 67340), we find a draft of a donation contract with numerous interlinear insertions of words and phrases, which seem to have been added by Dioscorus himself (P.Cair.Masp. 3, p. 165). Obviously, his wills, donation and marriage contracts proceeded through various stages of composition, characterised by different types of corrections.

4.3 Petitions and drafts

The high number of corrections per text (Fig. 2), especially at lexical and phrasal levels (Fig. 4), could point towards the presence of drafts among petitions in the papyrological corpus. A good example of a draft are the petitions in the archive of Aurelius Ammon *scholasticus*, son of Petearbeschinis, dated to the fourth

century CE (TM ArchID 31; see also Luiselli 2010: 82).[14] The archive preserves documents from a wealthy and educated family of Egyptian priests in Panopolis (cf. P.Ammon 2, pp. 21–2). Most of the petitions in the archive are related to one court case concerning the ownership of domestic slaves after the death of Ammon's brother Harpokration in 348 CE (P.Ammon 2, 32–46). At least eleven papyrus sheets have been used for drafts of Ammon's petition(s) addressed to the *katholikos* and the prefect of Egypt, in which he tries to prove that he is the rightful heir of his brother (see P.Ammon 2, pp. 11–21). The situation is introduced as follows in two successive drafts of the petition, written by Ammon himself:

> (1) P.Ammon 2, 41, 16–19
> πρ[ὸ] πολλοῦ τ[ινος χρόνου \ὸ/ ἀ]δελφός ⟦τις ἐμὸς⟧ \μου/ | Ἀρποκρατίων τοῦ[νομ]α̣ π[ε]ρ̣ὶ λόγους καὶ αὐτὸς [ἐ]σπουδακὼς ἀποδημ[ίαν ὑ]περόριον ἔξω τῆς | Αἰγύπτου τυγχάνει σ̣τ̣ειλάμενος. ἐν̣τ̣εῦθεν δὲ ἀποδημῶν ἀπὸ τῆς λα̣μ̣πρᾶς ταυτησὶ πόλεως | ⟦κατέλειπεν ἀνδράποδα ἑαυτοῦ ἐνταυθὶ⟧ κατέλειπεν ἀνδράποδα π[α]ρ̣' ἐμ[ο]ὶ̣ ἐνταυθοῖ τότε διατρίβ[ο]ντι
> 'For some long time ~~a brother of mine~~ my brother, named Harpokration, who also studied rhetoric himself, was preparing a journey abroad outside of Egypt. When he departed then from this illustrious city here (i.e. Alexandria), ~~he left his slaves here~~ he left slaves at my place, because I then resided here.'

> (2) P.Ammon 2, 45, 1–4
> [πρὸ πολλοῦ] τινος χρόνου οὗτο[ς ὁ ἀδελφός] μ[ου] Ἀρποκρατίων τοὔνομα περὶ λόγους | [καὶ αὐτὸς ἐσπουδακὼς ἀποδη]μ̣[ία]ν̣ ἔξω τ̣[ῆς Αἰγύπτ]ου τυγ[χάνει] | στειλάμενος. ἐντεῦθεν δὲ ἀ̣π̣ο̣δημῶν κατέ[λε]ι̣πεν ἀνδράποδα ἑαυτοῦ | ἐν τῆιδε τῆι πόλει προδιατρίψα̣[ντι τότε] ἐ̣[μο]ὶ̣
> 'For some long time my brother, named Harpokration, who also studied rhetoric himself, was preparing a journey outside of Egypt. When he departed then, he left his slaves in this city (i.e. Alexandria), because I then resided here.'

P.Ammon 2, 45 is written in Ammon's formal hand and considered to be the last version of the drafts preserved (see P.Ammon 2, pp. 43–50), while P.Ammon 2, 41 is found on the *verso* and in the margins of the *recto* of P.Ammon 2, 30 and

---

[14] Although the title *scholasticus* is often used by juridical experts, it does not refer to a profession as a notary or lawyer, strictly speaking, but rather to a generally high level of education in grammar, rhetoric, philosophy and literature, cf. P.Ammon 2, pp. 21–2, and references there.

written in Ammon's fast 'drafting hand' with numerous corrections. For example, ἀ]δελφός ⟦τις ἐμὸς⟧ ('a brother of mine') seems to have been replaced here by \ὁ/ ἀ]δελφός \μου/ ('my brother'), which is also adopted in later versions. The beginning of line 19 shows that changes in composition were often made during writing: the first words of the line ⟦κατέλειπεν ἀνδράποδα ἑαυτοῦ ἐνταυθὶ⟧ ('he left his slaves here') seem to have been deleted and replaced by the following κατέλειπεν ἀνδράποδα π[α]ρ̣' ἐμ[ο]ὶ̣ ἐνταυθοῖ ('he left slaves at my place here'). Many additional changes without a precedent in the previous versions have found their way into 45, such as the omissions of the superfluous ὑ]περόριον ('abroad') and ἀπὸ τῆς λαμπρᾶς ταυτησὶ πόλεως ('from this illustrious city here'), and the changes to the construction at the end of the phrase. Of course, changes in composition do not always have to be indicated by deletions and insertions in the text itself, they can also have be introduced without explicit mention from one version into the other or in additional drafts that have not been preserved to us.

The revisions found in the fifteen (parts of) petitions published as P.Ammon 2, 32–46 concern predominantly changes in formulation, such as word choice and syntax (see examples 1–2 above). Most of Ammon's corrections, therefore, are deletions and insertions of words (45%) and short phrases (40%) rather than changes to graphemes or morphemes (15%). Especially in the parts casually penned down in the margins, Ammon is continuously searching for improvements in the formulation of the message, e.g. by rephrasing ⟦[καίπ]ερ οὔτε⟧ πεποίηκε̣ν ('although nor did he do') to \ἀλλ]ὰ οὐδὲ τ[ο]ῦτο/ πεποίηκε̣ν ('but also this he did not do') in 41, 68 and the replacement of ⟦ἤδη⟧ by \λοιπὸν/ ('already') in 41, 70. These changes and additions to his own words are likely to have been made immediately after finishing the phrase or perhaps upon rereading a sentence or section. It would have been more difficult to review the text as a whole, because by that time the different parts of 41 would have been spread out across the *verso* of the sheet and squeezed into the vertical and horizontal margins of another text on the *recto*. This type of extensive revision at a lexical and phrasal level during writing seems typical for the preliminary stages of the composition process commonly associated with drafts.

Although petitions may often preserve corrections as result of a drafting process, not all corrected petitions are drafts. While drafts of petitions are left behind in the (private) archives of scribes or thrown away, the final versions are sent off to the authorities in the district capitals and Alexandria. For example, the minor corrections in the petitions addressed to and received by Apollonios in his function as the governor of Apollonopolites Heptakomias are, in fact, very similar in nature to the ones identified in the final versions of (private) letters addressed

to him (see Section 4.1). A group of petitions from Euhemeria also seems to have arrived at the authorities and form part of an official archive (TM ArchID 187). All documents have the same measures and have been produced by four or five scribes between 28 and 42 CE (P.Ryl. 2, p. 117). Most of them are addressed to the chief of the police (*epistates phylakiton*) or the governor of the Arsinoite nome, but all are assumed to have been copied and forwarded to the police chief of the village (*archephodos*) of Euhemeria (see P.Ryl. 2, pp. 117–19; Sijpesteijn 1989; 1992).

The petitions found in the office of the police in Euhemeria are thus forwarded copies of the final submitted versions, probably produced by professional scribes. Still, a considerable number (20 out of 33) of them contains at least one correction.[15] That these corrections are different from the ones in the drafts of petitions described above can be observed from their lower frequency (on average 2.4 per corrected text) and especially the linguistic level of the corrections. The absence of phrasal revisions in these copies is consistent with the production process. The composition of the text was already completed and the formulation did not need to be altered in any way during reproduction. Some of these corrections seem to remedy typical copy mistakes, such as Ὀρσεῦς corrected from Ορσενο mistaken for Ὀρσενοῦφις the line below (P.Ryl. 2, 149, 15), the ηρ of Ἡρᾶτος corrected from καί which happens to be the next word (P.Ryl. 2, 149, 16), and the anticipated δε corrected into ου in οὐσία[ς Δεκίμου] ('estate of Decimus') (SB 20, 15032, 5). The scribe of P.Ryl. 2, 142, 21 writes δεσχῶ(ν) instead of δεσμῶν ('bundles (of hay)'). The confusion between χ and μ seems difficult to explain phonologically or semantically, but could have been caused by visual copying (see also Yuen-Collingridge and Choat 2012). There are also various examples of words written in dittography in these petitions (without correction), e.g. τοῦ written both at the end of l. 6 and the beginning of l. 7 (P.Ryl. 2, 124), a double abbreviation for δραχμαί (P.Ryl. 2, 127, 30) and δημόσια written at the end of l. 19 and beginning of l. 20 (P.Ryl. 2, 149). Omission and repetition of words (especially at line breaks) are common features of copying from an exemplar. Although the copyists of these petitions may not have been extremely careful while copying, they did seem to consider it worthwhile to correct their mistakes.[16] The corrections to the copies of petitions

---

[15] Unfortunately, there are no images available for the majority of the petitions from this archive. My observations are based on the corrections as described by the editors in the editions of P.Ryl. 2, 124, 127, 142 and 149 and SB 20, 15032 (see Sijpesteijn 1989).

[16] There are other orthographic variants that this group of petitions have in common, e.g. sixteen

from Euhemeria have not been made during the process of composition of the text, as the lexical and phrasal corrections in the drafts of Ammon, but they have been added during production of the copied version. Some of the copying errors were immediately corrected, when noticed, usually by adapting or writing over the letter(s), cf. e.g. SB 20, 15032, l. 5 with note (Sijpesteijn 1989: 196).

Copying with self-correction generally results in corrections of different kinds of linguistic features than the process of stylistic revision during the composition of a text. On the other hand, we should allow for some overlap between different production processes. Although phrasal corrections are rare in copied texts, copy mistakes of more than one word may occasionally occur. The drafting process itself is likely to involve some copying as well: from one draft to the other or from the draft to the final version. This could explain some of the immediate corrections affecting smaller elements in Ammon's drafts. For example, the correction of the ε into τ in ἕως δὲ ⟦ε⟧ταύτην ἔτι ἐν χερσὶν εἶχον τὴν φροντί(δα) ('while I still had that concern in my hands') (P.Ammon 2, 41, 42) may at first seem to have been part of his stylistic revisions by introducing the demonstrative ταύτην to the phrase. It is more likely, however, that this is a correction of a copying mistake, since the demonstrative seems to have been present already in earlier versions of the petition (e.g. P.Ammon 2, 38, 28, and 39, c 9). The same phenomenon can be observed in τὴν ⟦δεσ⟧τούτων δε[σποτ]ε[ί]αν ('the ownership of these (slaves)') (P.Ammon 2, 45, 22). Again, the sudden introduction of τούτων is most likely a reaction to accidentally skipping this word, which has already been used in previous versions in this phrase (e.g. P.Ammon 2, 32, 18; 36, 6; 40, 24; 41, 47). This shows that the type of correction may not only give an idea about the stage in the process of composition, but also about the method(s) of production of the text in question.

---

out of the thirty-three petitions write ἀξιῶι instead of ἀξιῶ. Hypercorrection of the ι adscript (see also Vierros 2012a: 121–36) in other words than ἀξιῶ seems particularly common in 131 and 139, e.g. κώιμην for κώμην in 131, 14 and 139, 18, and these two texts happen to be written in the same 'stiff clear hand' according to the editor (P.Ryl. 2, pp. 127 and 136). Several other nonstandard spellings occur in smaller numbers of examples, such as ἑατοῦ for ἑαυτοῦ and ἀκθῆναι for ἀχθῆναι. Variant spellings could have been introduced accidentally or on purpose by the same copyist but could also have been present already in the exemplar unknown to us. P.Ryl. 2, 124 and 135 seem to have more problems involving the spelling and choice of morphemes than the other petitions in this archive, e.g. λιστρικο τρόπο εἰς ἃς γεορ|γο instead of λῃστρικῷ τρόπῳ εἰς ἃς γεωρ|γῶ ('in a thievish way to what I cultivate') (P.Ryl 2, 135, 7–8) and τῆς γυναι|κός μου Ἀπλουνοῦ|τος καὶ η (l. τῆς) ταύ|τ[ης] μητὴρ (l. μητρὸς) Θερ|τος ('my wife Aplounous and her mother Thermis(?)') (P.Ryl 2, 124, 7–11). It seems unlikely that all of these forms have been introduced later through careless copying or in a false attempt by the copyist to improve the language of the exemplar.

4.4 Lists under composition

Corrections in lists are less common than in the other three genres (Fig. 2) and the far majority are found at lexical (53%) and phrasal (33%) levels (Fig. 4). The almost complete absence of corrections to morphemes (average 3%) may be caused by the repetitive nature and general lack of syntactic complexity in lists of names or items. In the lists and accounts of Dioscorus, half of the corrections are insertions of letters and words above the line. This type of insertions is also very common in drafts of petitions and contracts, but in some lists (short) phrasal insertions are more frequent than insertions of single words. In P.Cair.Masp. 2, 67143, an account of people to be accused and a list of stolen animals (see Ruffini 2008: 161–3), Dioscorus adds after an amount of oil also \ἐραι(γμοῦ) φακ(ῶν) (ἀρτάβη), πριστ(ήρ)/ ('an *artabe* of pounded lentils, a saw') to the entry of the accused Hermaos (l. 18; see Youtie 1979: 96) and \ἄλλο αἰγί(διον) α/ ('one other kid') to 'three sheep that have been found' in the list of stolen animals (l. 23). To the list of names on the *verso* he adds \(καὶ) τοῦ διακ(όνου) ἄπα Μηνᾶ/ 'and by the deacon Apa Menas' (l. 12). Letters that have been forgotten are supplemented as well during the writing, such as the ε added above the line in Θερ\έ/σο(υ) (l. 25). In order to produce such a list of people and stolen goods, Dioscorus may have been collecting names and items from elsewhere to organise them into the three separate lists on this papyrus. During this process, he could easily have come across some extra information about the same individual which needed to be added to one of the previous entries. The formulation as well as the slightly larger, more hastily written letters in a darker shade of ink suggest that the phrase ὁμοί(ως) αἰγί(δια) β το(ῦ) (αὐτοῦ) Μ[ακα]ρ(ίου) ('similarly two kids by the same Makarios') was also added at a later stage to the entry of Makarios (l. 25).

The process of composition is also reflected in the corrections in lists in the archive of Petaus, village scribe of Ptolemais Hormou and surrounding villages (TM ArchID 182). Deletions are particularly common in these lists. In a list of names of persons from the village of Syron (P.Petaus 100), whole entries have been crossed out (l. 44), encircled (ll. 39–40) or wiped out (ll. 21–2 and 28), some more successfully than others (cf. ll. 1–3, 23–4). New entries have been added in between the lines (ll. 10, 43, 45, 52). The deletion of ⟦Ἀμμλῆς υἱὸς Πααῦ Κιασῶς⟧ ('Ammles, son of Paaus Kiasis') (l. 44) and subsequent insertion of Τασωίκιος ἐπ(ικαλούμενος) Ἀμλῆ ('Tasokis, nicknamed Amle') (l. 45) may have been prompted by a confusion of both persons named Amles (see *ed. pr.*, n. to ll. 43–5). Most of the remaining entries on the *recto* have been marked with an 'x' in the left margin. The deletions, insertions and marking of the entries on this

papyrus indicate stages in a selection process. P.Petaus 59 preserves several drafts and a copy of the final version of a list of individuals proposed for the liturgical function of *sitologos*. Deleted names in the first draft (f; between ll. 24–5, 32–3 and 33–4) are left out in the subsequent versions, while added metronymics (ll. 27–8) become included in later ones. The empty space meant for the prescript at the top of the later versions (a–d) betrays several failed attempts to move towards a more finalised version of the list, as the editors conclude 'Alle Urkunden zusammen zeigen jedoch eindringlich, ein wie mühsames Geschäft es war, eine solche Vorschlagsliste aufzustellen.' (P.Petaus, pp. 230–31). It is clear that the main aim of these lexical and phrasal corrections in lists is not the improvement of the language, but the improvement of the contents.

 Just as the drafting process of petitions includes not only changes to the formulation and textual composition (see Section 4.3), the composition of a list also yields more than just changes to the content. As the multiple versions of P.Petaus 59 show, the process of composition in various stages means that copying between drafts, from draft(s) to final version and perhaps from final version to multiple copies can be part of the production of a list. The patronymic of Ψονθ(νεῦς) Παθύνεως in the draft versions (a), (b) and (c) was copied into the draft version (d) as Ψαθύνεως (l. 53), while the copy of the final version erroneously duplicates the patronymic Ἀπύγ(χεως) of the line before in its place. The same error is made in ll. 36–7: the copy of the final version interchanges the order of the two entries and writes τοῦ Τεσενούφεως to both Ἥρων and Φιλάμμ(ων) Νεσεῦτ(ος), while the previous drafts (a, b, d, e and f) give τοῦ Φιλάμμωνος as the grandfather of Philammon.[17] Copy mistakes can also elicit corrections, when they are discovered during writing or afterwards. In a list of names arranged by families living in Ptolemais Hormou (P.Petaus 93), corrections are usually added immediately. For example, the patronymic Ἰσχυρίωνος was initially skipped in Κεφαλᾶς ⟦Κεφαλ⟧ Ἰσχυρίωνος τοῦ Κεφαλᾶ ἀφῆλ(ιξ) (ll. 124–5), but the anticipated grandfather's name Κεφαλᾶ was encircled as soon as the mistake was discovered half–way through writing the λ. Similarly, the patronymic Ἰσχυρᾶ was initially omitted in Ἰσχυρᾶς \Ἰσχυρᾶ/ ἐπικαλού(μενος) Κορ[κό]δ[ιλο]ς (l. 75) and added as soon as the mistake was noticed, probably between writing ἐπικα and λού(μενος) causing a small space between the two parts of the word

---

[17] The copy of the final version of 59 seems to have been produced hastily and contains multiple copy mistakes (cf. *ed. pr.*, p. 230), such as the nickname of Ἀπύγχ(ις) Πααῦτ(ος) written as υἱὸς Τῖνος (l. 48) instead of Οὐιστῖνος (a–b) or Οὐστῖνος (d) for the Latin name Vestinus, perhaps caused by confusion with the previous entry of Ἀπύγ(χις) Παθ(ύνεως) (also Πααῦτ(ος) in a and d) ἐπικ(αλούμενος) υἱὸς Μούιτ(ος) (l. 47).

(see *ed. pr.*, n. to l. 75). Comparable corrections of skipped letters and words are also found in the list of names in P.Petaus 100 (see also above). For example, the π (anticipating the patronymic) was immediately corrected into the υι of υἱός in Πααῦς υἱὸς Πετειμέννις (l. 15), and cf. also the υἱός skipped and added later by the same scribe in Ποτάμων \υἱὸς/ Κολμοῦς (l. 26). These types of small scribal errors and copy mistakes are an intrinsic part of the production process of multiple versions of any document and lie also behind some of the corrections made in lists.

5 Conclusions

Scribal corrections commonly occur in documentary papyri. On average, they are more likely to be found in petitions (32%) and letters (28%) than in contracts (22%) and lists (19%). Corrections in petitions and lists tend to come in higher numbers per text and the majority of the corrections are concerned with words and phrases, while the majority of the corrections in contracts affect graphemes and morphemes. These differences between the genres could be related to the structural properties and function of the text as well as the production process of the document and the stage of composition that is preserved to us. Based on qualitative analysis of corrections in various archives, it is possible to distinguish two basic methods of composition (free composition and copying) and two main stages of production (preliminary and final version). Different methods may coincide at various stages in the production process.

The private letters in the archive of the governor Apollonios illustrated that final versions of letters may contain small numbers of corrections at various levels. Accidental scribal errors, nonstandard orthography and morphological endings are corrected immediately or upon rereading the text. Second thoughts on the formulation and/or content of the message may be responsible for lexical or phrasal revisions during writing. In letters, the preliminary and final stages of composition often coincide. Final versions of contracts attract similar corrections of graphemes and morphemes, but lack revision of formulation and contents, probably because extensive revision of semantics and syntax was legally unacceptable in final versions of juridical documents. The non-final versions of contracts in the archives of the Alexandrian office and the scribal office of Kronion, on the other hand, show that many contracts are at least partially composed by a scribe. These preliminary versions of contracts are characterised by changes at lexical and phrasal levels, just as the frequent deletions and insertions

of words and short phrases found in drafts of petitions. Documents produced through several stages of drafting usually involve copying as well. Copy mistakes may occur between drafts, from draft to final version or in copies of the final document, as visible in the series of drafts by Ammon and the lists produced in the office of Petaus. Lexical and phrasal corrections in drafts of lists usually reflect changes to the content rather than changes in formulation.

Since many documents contain a limited number of corrections, the presence or absence of corrections itself is usually not enough to distinguish between a preliminary and final version of a document. The linguistic level of the corrections seems to provide a more informative criterion. Scribal revisions in all four genres confirm that lexical and phrasal corrections are typically found in documents at preliminary stages of (free) composition, while corrections of graphemes, morphemes and (parts of) words are also encountered in final versions and as a result of copying. Although both methods of production can be applied at preliminary and final stages of composition, the types and linguistic levels of scribal corrections could provide a helpful tool for identifying scribal practices at different stages in the process of textual production.

References

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: 229–47.

Yuen-Collingridge, Rachel and Choat, Malcolm 2012. 'The copyist at work: scribal practice in duplicate documents', in Paul Schubert (ed.), *Actes du 26e Congrès international de papyrologie: Genève, 16-21 août 2010*. Geneva: Librairie Droz S.A, pp. 827–34.

Depauw, Mark and Stolk, Joanne Vera 2015. 'Linguistic Variation in Greek Papyri: Towards a New Quantitative tool for Study', *Greek, Roman and Byzantine Studies* 55: 196–220.

Gignac, Francis Thomas 1976. *A Grammar of the Greek Papyri of the Roman and Byzantine Periods. Volume I: Phonology*. Milan: La Goliardica.

Gignac, Francis Thomas 1981. *A Grammar of the Greek Papyri of the Roman and Byzantine Periods. Volume II: Morphology*. Milan: La Goliardica.

Homann, Margit 2012. 'Eine Randerscheinung des Papyrusbriefes: der *versiculus transversus*', *Archiv für Papyrusforschung* 58: 67–80.

Leiwo, Martti 2017. 'Confusion of Moods in Greek Private Letters of Roman Egypt?' in Klaas Bentein, Mark Janse and Jorie Soltic (eds.), *Variation and Change in Ancient Greek Tense, Aspect and Modality*. Leiden – Boston: Brill, pp. 242–60.

Luiselli, R. 2010. 'Authorial Revision of Linguistic Style in Greek Papyrus Letters and Petitions (AD i–iv)' in Trevor V. Evans and Dirk D. Obbink (eds.), *The Language of the Papyri*. Oxford University Press, pp. 71–96.

Papathomas, Amphilochios 2018. 'Das Ringen um korrekte Sprache, guten Stil und rechten Sinn. Grammatische und stilistische Verbesserungen auf spätantiken griechischen Papyrusbriefen (5.-8. Jh. n. Chr.)', in Pierre Swiggers (ed.), *Language, Grammar, and Erudition: From Antiquity to Modern Times. A Collection of Papers in Honour of Alfons Wouters*. Leuven – Paris – Bristol: Peeters, pp. 145–66.

Ruffini, Giovanni 2008. 'Factions and Social Distance in Sixth-Century Aphrodito' in Jean-Luc Fournet and Caroline Magdelaine (eds), *Les archives de Dioscore d'Aphrodité cent ans après leur découverte. Histoire et culture dans l'Egypte byzantine. Actes du Colloque de Strasbourg, 8-10 décembre 2005*. Paris: De Boccard, pp. 157–70.

Seidl, Erwin 1973. *Rechtsgeschichte Ägyptens als römische Provinz (Die Behauptung des ägyptischen Rechtes neben dem römischen)*. Sankt Augustin: H. Richarz.

Sijpesteijn, Pieter Johannes 1989. 'Another οὐσία of D.Valerius Asiaticus in Egypt', *Zeitschrift für Papyrologie und Epigraphik* 79: 194–6.

Sijpesteijn, Pieter Johannes 1992. 'Petition to the Chief of Police', *Zeitschrift für Papyrologie und Epigraphik* 91: 101–2.

Sijpsteijn, Pieter Johannes and Worp, Klaas Anthony 1977. 'A Latin Papyrus from the Vienna Papyrus collection', *Zeitschrift* für *Papyrologie und Epigraphik* 24: 91–4.

Stolk, Joanne Vera 2015. 'Scribal and phraseological variation in legal formulas: the adjectival participle of ὑπάρχω + dative or genitive pronoun', *Journal of Juristic Papyrology* 45: 255–90.

Stolk, Joanne Vera 2019. 'Itacism from Zenon to Dioscoros: scribal corrections of <ι> and <ει> in Greek papyri', in Alberto Nodar, Sofía Torallas Tovar, María Jésus Albarrán Martínez, Raquel Martín Hernández, Irene Pajón Leyra, José-Domingo Rodríguez Martín and Marco Antonio Santamar (eds.), *Proceedings of the 28th Congress of Papyrology, Barcelona 1-6 August 2016*. Barcelona: Publicacions de l'Abadia de Montserrat, pp. 690–97.

Stolk, Joanne Vera 2020. 'Orthographic variation and register in the corpus of Greek documentary papyri (300 BCE – 800 CE)', in Klaas Bentein and Mark Janse (eds.), *Varieties of Post-Classical and Byzantine Greek*. Berlin: De Gruyter Mouton, pp. 299–326.

Van Beek, Bart and Depauw, Mark 2013. 'Quantifying Imprecisely Dated Sources: A New Inclusive Method for Charting Diachronic Change in Graeco-Roman Egypt', *Ancient Society* 43: 101–14.

van Groningen, Bernard A. 1932. 'Projet d'unification des systèmes de signes critiques', *Chronique d'Égypte* 7: 262–9.

Vierros, Marja 2012a. *Bilingual Notaries in Hellenistic Egypt. A Study of Greek as a Second Language*. Brussel: KVAB Press.

Vierros, Marja 2012b. 'Phraseological Variation in the Agoranomic Contracts from Pathyris', in Martti Leiwo, Hilla Halla-Aho, and Marja Vierros (eds.), *Variation and Change in Greek and Latin.* Helsinki 2012, pp. 43–56.

Youtie, Herbert C. 1979. 'P. Mich. I 2: ἐρεγμὸς φακῶν', *Zeitschrift für Papyrologie und Epigraphik* 34: 95–6.

3. Whose Words? Identifying Authors in Greek Papyrus Texts Using Machine Learning

Marja Vierros & Erik Henriksson

1 Introduction[1]

Imagine a situation where Ptolemaios, son of a Greek soldier-settler Glaukias, is compiling a letter to his brothers in Memphis, Egypt. He has learned to read and write Greek (an advantage of his background), but he does not write very often. He picks up the reed-pen (*kalamos*) and finds a piece of papyrus (perhaps by cutting it from a larger old scroll with Demotic Egyptian writing and washing the old writing away). He then composes his message, writes it on the papyrus sheet and sends it away to the addressees. His brothers read and understand the message despite the spelling mistakes of Ptolemaios, whom they know to be an unaccustomed writer. This message of Ptolemaios is preserved to the modern day (UPZ 1.67) along with many other texts he, his brother Apollonios, and their companions in the Memphis Sarapieion wrote. We can thus read what Ptolemaios wrote in his own, recognised, hand in 152/3 BCE and we can interpret his message as his own language production. We can also assume that his spelling mistakes mirror, firstly, his education in writing, his lack of practice, and his way of speaking, and, secondly, they can also reflect how other Greek speaking people around him communicated.

Here is another situation to imagine: the same Ptolemaios has suffered from an action he felt was wrong towards him and towards the twin girls he was a custodian of, and he approaches the Greek legal officials to seek justice. Ptolemaios might think that his writing skills are not good enough to produce an effective petition. He consults his youngest brother Apollonios, who has been educating himself in the art of writing petitions to officials by copying such documents produced by educated scribes. Ptolemaios is narrating the situation and Apollonios is writing it down. They draft a petition in which the narrative, choice of words, and syntax primarily come

from Ptolemaios, though Apollonios also influences the text's composition, since he knows some of the phrases that are customary to the petition genre. The final version of the text is certainly produced by a professional scribe, who recomposed the text drafted by Ptolemaios and Apollonios, polishing the language and the phraseology. In this case, we perhaps take this petition as reflecting the language of the scribe (or Apollonios) and not that of Ptolemaios, even though his name is mentioned as the sender of the petition, whereas the name of the scribe is not mentioned anywhere. In this Chapter, we will use the term *author* for the role of Ptolemaios and the term *writer* for the role of the scribe. In the first case presented above, Ptolemaios himself was both the writer and the author.

The documents providing evidence of the processes described above have survived in the so-called archive of the Katochoi in the Memphis Sarapieion.[2] They lead us to the general problems we have when using documentary papyri and ostraca as linguistic data, especially when we want to apply sociolinguistic analysis: what was the actual number of people behind the text we read? Are we analysing the language produced by the person who penned the letters on the surface of the papyrus or potsherd, or are we analysing a product of an author who compiled and drafted the text, but did not actually write the final product (however short it was). Or are we analysing a merger of two or more persons involved in the production: the actual writer and the author(s) who drafted or dictated the contents? Obviously, the writer and the author could have been one and the same person, but earlier models or templates could also be followed or stock formulae used. If the writer was not in charge of the content production, he perhaps made his own fingerprint visible unconsciously at some levels of the language, e.g. in the orthography (the most obvious option) or in producing inflected forms, i.e. morphology, differently from what the author had meant, or he could even have had an impact on the syntax if the draft was not detailed enough.

Why is it important to differentiate these roles? The distinction becomes meaningful, for instance, when we are trying to trace idiolectic usage versus commonly shared features in language use. The Greek papyrological sources from Egypt are preserved unevenly and by chance, meaning that some preserved archives include a large quantity of texts authored by the same people and, therefore, some idiolectic variation attested can distort the image of overall

---

[2] Published in the *UPZ I* edition (Wilcken 1927). As for the petitions, there are several different groups of drafts, copies and final versions concerning different situations in the life of Ptolemaios and his brother Apollonios, many of them concerning two girls under Ptolemaios's protection, Taues and Taes, who were acting as 'twins' in the temple cult. The language of the archive has been discussed, e.g. by Bentein (2015) and Vierros (2020).

language use, if we do not know that one and the same person was behind the variation. With the archives especially, we have knowledge of some of the authors involved: we may know their names, family, age, gender, ethnicity, language skills and education, but usually we have none of those facts for the scribes; if we do not know which texts the agents wrote themselves and which they just had some 'author influence' upon, we cannot use the linguistic result for sociolinguistic analysis or conclusions.

Therefore, it is important to determine if it is possible to differentiate which types of linguistic variation can be attributed to a writer who did not necessarily compose the text, and which to an author who did not necessarily write the text down themselves. For example, sometimes there are copies, and in those the actual copyist probably did not have much influence on the result (e.g., Vierros 2019). On the other hand, some drafts have survived in which certain linguistic levels are affected by the author and some by the writer. Thus, it is interesting to see whether the authorship attribution and author profiling methods can identify these differences.

In this chapter, we will take first steps towards using authorship attribution methods with short and fragmentary Greek documentary papyri from the Hellenistic period (3rd to 1st centuries BCE). Our data have authors identified and marked as metadata, so it is possible to test if the methods work in a way that can help later with other data of unknown authors. Text clustering can also tell us something about language varieties in general; for example, how much formulaic phraseology affects the results. We will first briefly introduce some background on authorship attribution methods, also known as stylometric analysis. The topic has been covered far and wide in several studies, so we will only deal with the tip of the iceberg suitable for the analysis of ancient material.[3] Then, we will present the methods used in this preliminary study. In Section four, the test data selected for this Chapter is introduced and the test results are discussed. Finally, we conclude what we have learned from these test cases.

2 Authorship attribution and short, fragmentary texts

With ancient Greek documentary papyri, there are several characteristics to consider. First, the texts are often short and, second, they are fragmentary, i.e. they may have been preserved only partially (different amounts of information being missing:

---

[3] A good survey is provided by Stamatatos (2009) and updated in Stamatatos (2016).

parts of words, parts of sentences, several sentences). Third, several text types differ from one another in their language use: private letters differ from business letters in topics and level of colloquiality; administrative texts, like petitions to authorities, have their own phraseology and a narrative part; contracts have set formulae as well as other register-related features. Only occasionally is it possible to assume a single writer or author behind several documents through text-external factors. That is the case when the handwriting can be identified and combined with the contextual information a group of texts can offer (the names of writers and addressees, the date and place of writing, etc.). Sometimes the archaeological context is also known, so we know that certain texts are related in some way.

Fortunately, the development in authorship attribution studies is progressing fast in connection to the rapid development of more powerful computers, algorithms and machine learning methods. Data handling is easier and we have more data in electronic form, so many features can be studied and combined more efficiently than before; especially when a text can be morphosyntactically tagged (automatically or by other means), there are plenty of features that can be counted and analysed. In this Section, we will first give a short summary of authorship attribution methods and their development, and then turn to those issues that seem most useful with primary historical sources.

2.1 Authorship attribution in a nutshell

Authorship attribution is a specific field in computational science and is summarised by Stamatatos (2009: 538) in the following way: 'The main idea behind statistically or computationally-supported authorship attribution is that by measuring some textual features we can distinguish between texts written by different authors.' The problem of finding out the true author of a text comes up e.g. in forensic linguistics, and the fundamental task of authorship attribution deriving from real world situations has been described by Koppel et al. (2012: 284): 'given two (possibly short) documents, determine if they were written by a single author or not'. A sub-species of open-set authorship attribution, called authorship verification, is an attribution problem where, given a set of texts by the same author, the task is to determine whether a text under investigation is by that author or not (Stamatatos 2016: 10).

Authorship attribution is continuation of *stylometry*, a method of counting features that can distinguish differences between writing styles.[4] These features can

---

[4] In addition to this brief historical recap, a survey of the methods and features used in authorship

be, e.g., sentence length, word length, word frequencies, character frequencies, vocabulary richness, etc. These studies have a long history. Writing style was being quantified long before the age of computers; one of the earliest stylometric studies concerned the identity of the author of the plays written in Shakespeare's name (Mendenhall 1887). A seminal work in the field of modern authorship attribution was Mosteller and Wallace's Federalist Papers study (1963), which was performed to test the methods and Bayesian statistical analysis based on probability calculus. Mosteller and Wallace ended up using a small set of very common words (e.g., 'upon', 'an' and 'to'), which produced significant discrimination results between the candidate authors.[5]

The basic idea is that writers use a set of features in a way that makes their style so distinct that it is almost like a handwriting or fingerprint. The feature combination may differ within one writer if the text type and target audience are different; consequently, unconscious features are the best for this type of analysis. This is what is known as the *stylome*.[6] Therefore, analysing the way *function words* (i.e., words that carry no semantic information, usually articles, prepositions, pronouns, etc.) are used has kept its value, because 'function words are used in a largely unconscious manner by the authors and they are topic-independent. Thus they are able to capture pure stylistic choices of the authors across different topics'.[7] However, it has been found that combining several different methods may help further, especially with short texts (see below).

---

attribution can be found in, e.g., Grieve (2007), Stamatatos (2009), or, with a focus on ancient texts, Gorman & Gorman (2016) and Pavlopoulos & Konstantinidou (2023).

[5] They found that average sentence length, for example, was not helpful in identifying the author, as all three wrote in an oratorial and overwhelming style which was common in their time. In addition to common words, there were certain lexical minimal pairs that had been noticed earlier: H used 'while', whereas M used 'whilst'. However, these words had relatively low frequency (per 1,000 words), so there could be a long passage of text where these words never appeared; therefore, the frequent word lists were necessary as well.

[6] Van Halteren et al. (2005); they start from the idea that each person learns their own variant of their native language that differs from those used by other speakers of the same language. The question in their article is whether this can be measured and quantified to identify different speakers. The study is performed by examining written texts (by Dutch students) via combined vocabulary and syntax features, and the answer is clear: a measurable *stylome* exists. Syntax is assumed to be less distinctive when compared to vocabulary; apparently, by the age of eight, syntactic features are already quite fixed but vocabulary grows with age.

[7] Stamatatos (2009: 540). It is not necessarily easy to draw the line where words like pronouns or prepositions begin to have enough semantic value to be considered content words. Stamatatos (loc. cit.) lists several studies and their selections of function words.

2.2 Short texts – syntax and idiosyncrasies to the rescue

Documentary papyri come in different lengths, but a very common feature they share is shortness. Some very long texts do survive, like P.Petra 4.39, of around 5,000 words, but these are rare. The shortest documents that are still considered to be one 'text' can consist of fewer than ten words (e.g., O. Abu Mina 1). The most common range of length for letters, contracts and petitions is, in our estimation, 50–300 words per document. Moreover, the texts have gaps due to the poor survival of the material, and contain abbreviated words as well as words that have been difficult or impossible to read.

Several studies have taken on solving this pertinent problem of authorship attribution, since real-life situations often also deal with samples that are small. Maciej Eder (2015) tackled the issue of sample size – that is, how many words would be the minimum for positive results in authorship attribution. He found that it varied somewhat depending on the language and genre; Latin prose required only 2,500 words, but poetic corpora in Greek, Latin, and English required some 3,000 or more. He concluded that, irrespective of language and genre, secure results require a minimal sample of 5,000 words (Eder 2015: 180). Eder compared different methods and features; interestingly, the best results were gained by analysing the 200 most frequent words (MFW) selected randomly from the corpus (the so-called bag-of-words method).[8]

Many studies examining short texts have concluded that a good size of training material is significant. For example, newspaper columns of 300 to 5,000 words could be identified from 50 authors, but training data consisted of a minimum of 10,000 words per author on varied topics (Sanderson & Guenter 2006). We will rarely find 10,000 words per known writer or author in papyri. Another study attributed short text blocks of 1,000/500/200 words of Charlotte vs Anne Brontë, but overall, the texts from which the blocks came were long (250,000 words).[9]

The question of how similar a style closely related people with similar education can use is present also in the study on the Ciceronian corpus and

---

[8] Some improvement was gained by combining part-of-speech tags or character n-grams with MFW.

[9] Hirst & Feiguina (2007); noteworthy is that they utilised, e.g., syntactic parsing, part-of-speech-tags and bigrams of these. As a result, 1,000-word blocks could be attributed with 99.5% accuracy using the set of different syntactic features, but when block size was smaller, accuracy also decreased (500 words: 94.2%, 200 words: 87.5%). The result improved for the smaller blocks when additional lexical features were measured (96.8% and 92.4%).

especially the text *Rhetorica ad Herennium* (Vainio et al. 2019), which aimed to establish whether that text was written by Cicero or not (as had been suggested already in the fifteenth century). The Ciceronian corpus is large: 900,000 words. Vainio et al. used two different classifications: (1) word unigrams and bigrams (one word and two consecutive words) and (2) character five-grams (five consecutive characters). With the first, the influence of the text's topic is reduced by masking the content words and reducing them only to their part-of-speech tags; this helped in focusing on the style of the writer. Two classifiers were used (Support Vector Machine and Convolutional Neural Network). Interestingly, they found that a text usually attributed to Cicero's brother, Quintus, was in fact largely written by Marcus Tullius Cicero himself (*Commentariolum petitionis*),[10] but for *Rhetorica ad Herennium*, no certain conclusions could be drawn by the computational methods presented in the article.

A more recent study by Manousakis and Stamatatos (2023) on Aeschylus's *Seven Against Thebes* was concerned only with the end of the play and differentiated between sections of only a few lines, but the comparative material consisted of five complete secure plays by Aeschylus, Sophocles, Euripides, and Aristophanes, respectively. The methods in this study were all based on character n-grams.

One aspect that deserves mention in our case is the measurement of writing errors and idiosyncrasies. This, in a way, combines the real-life, human-based authorship attribution with the computer-based analysis. It is assumed that some idiosyncratic usages serve as unique fingerprints of a given author. This is especially the case in *unedited* texts, which immediately sounds familiar to papyrologists, particularly regarding private letters. These seem to be closer to what can be understood as unedited text, whereas in the scribal sphere, professional education usually prevented scribes from writing whatever they heard from their clients. The editorial filter of the scribes was, of course, not uniform in all cases; there was a wide continuum in levels of writing education, so even the scribes could produce their own idiosyncrasies. For example, a certain idiosyncratic relative clause structure was unique to one particular notary in Hellenistic Pathyris.[11]

---

[10] It was considered highly unlikely that brothers, even with similar education and background, could write in such a similar style that the computational methods would not detect this (only four short letters, c. 440 words, survive by the brother Quintus, so this could not be computationally tested) (Vainio et al. 2019: 35).

[11] Vierros (2012). See also Vierros (2020) on idiolectic spelling of Ptolemaios from the Katochoi archive. However, personal features could also be unstable; cf. Evans (2005), who discussed the omission of the valedictory closing word (ἔρρωσο) from the autograph letters by one Hierokles in the Zenon archive, and Evans (2010) on Amyntas's special aspirated perfect form that appears more

That feature was, in fact, a 'smoking gun', like the 'while/whilst' in the Federalist Papers (see above). Certain features could act as smoking guns, but in short texts they just might not be present.

The authorship attribution study by Schler & Koppel (2003) seems especially relevant in this respect, as they used short email discussions with an average length of 200 words as their data. They combined lexical features (function words), part-of-speech tags, and idiosyncratic usage (of various types: syntactic, formatting, spelling) and found that, especially when taking the idiosyncrasies into consideration, their methods worked well. However, counting only idiosyncrasies did not work.

In addition to idiosyncrasies, syntactic features have been used to some extent in several earlier studies mentioned above. Syntactic annotation's role was the main focus in Baayen et al. (1996); it was found to be truly useful, but only after first masking out some register-related features.[12] Gorman & Gorman (2016) investigated whether dependency syntax treebanks of Ancient Greek could be applied to questions of authorship in ancient Greek historiography; they extracted so-called syntaxWords from the treebanks that indicated the dependency paths, i.e., syntactic structures. They found similarities but also clear differences between the text of Polybius Book 1 (survived through direct transmission) and Books 9–10 (survived through excerptor). In a subsequent article, R. Gorman (2020) made use of the treebanked Ancient Greek texts to measure how short text passages could actually be verified for one author, and even with severe simplification of the syntactic strings, the results indicated that even 50-word sequences could be attributed to one author. He used a very large training set to achieve those results.

In sum, for short texts, one is encouraged to test with multiple features and methods, including function words, idiosyncrasies, and syntactic structures. One essential factor seems to be the register or genre; within one register, certain prominent elements may be more attributable to register than author, but once recognised, the remaining features may reveal author-specific style more easily. Another important factor when considering papyrological documents is the number of tokens from known authors to which we compare the unknown ones.

---

frequently in his autograph texts than elsewhere, but not in all of his texts.

[12] Baayen et al. (1996: 121–2) note that differences in register and text type are reflected in relative frequencies of linguistic variables, and thus there is a need to establish the range of variation of different writers within the same register or text type, as well as of the same writer in different registers or text types.

Our general token number for the whole corpus of Greek language is significant, but it consists of a large pool of different genres and registers, literary dialects, and more colloquial texts. For one register, the number of tokens may occasionally be very small.

3 Methods used in this study

Given all the challenges discussed above, we simply decided to study how certain word vector models and algorithms for author profiling and clustering would perform with documents whose authors (and writers) we already know. This way we can analyse whether the methods really point us to the author, or even the writer, and what other features emerge.

3.1 Data preparation and metadata

In the PapyGreek database,[13] we have marked up additional metadata for each annotated (treebanked) text and for some unannotated ones too, concerning the writers, authors, addressees, and external scribal personnel; they may all influence the language in which the papyrus has been written. First, the texts have been divided into *acts of writing* by utilising the Text Encoding Initiative (TEI) tag <handshift>[14] in order to study each different handwriting section as one language sample. Each act of writing has its WRITER = one person penning down the words on the papyrus; this writer may be identified as the same person writing on another papyrus. Such identifications are usually mentioned by the editors of the texts or by later scholars. Therefore, occasionally we can use the same ID for the writer in all the texts in which the same hand appears. This is a huge asset in the field of digital papyrology, since the writers are usually anonymous and do not appear in the digital prosopographic databases.

Each act of writing also has its AUTHOR, who was responsible for drawing up the text; i.e., the authors are the people who are usually named in the text itself as, for example, the person writing/sending a letter to someone, or the person who is the protagonist of a contract or a petition. The author could also be a notary who signed the document in their name. Quite often, we have more than

---

[13] See, e.g., Henriksson & Vierros (forthcoming) and Vierros & Henriksson (2021).

[14] The corpus of documentary papyri utilises the EpiDoc XML version of TEI (see https://epidoc.stoa.org).

one text preserved from the same author, especially in archives. We have given the authors their own PapyGreek ID number.[15] The author could be the same person as the writer, but we cannot usually assume this. We deal with the addressees in the same way; when they are named in the papyrus as addressees of the text, they receive a PapyGreek ID.

In marking the writers and authors, we have included the possibility to express uncertainty in identification. The uncertainty may arise from difficulties in handwriting identification or from fragmentarily preserved names in the text.

3.2 Authorship attribution and profiling methods

We examine authorship from two perspectives: 1) *author profiling* through document clustering and 2) *author attribution* through classification. In author profiling, we examine how texts naturally group together based on shared features, and in author attribution, we train a classifier to identify authors by learning from examples of their known texts.

In both types of analysis, we use four types of features: *character n-grams*, *function words*, *part-of-speech (POS) tags*, and *orthographic variations*. Character n-grams – which Manousakis & Stamatatos (2023) relied on exclusively in their attribution analysis of Aeschylus – represent sequences of *n* consecutive characters extracted from texts. For example, for the word *kalamos*, bigrams (*n*=2) would produce the sequence 'ka', 'al', 'la', 'am', 'mo', 'os', while trigrams (*n*=3) would give 'kal', 'ala', 'lam', 'amo', 'mos'. Our implementation can extract and combine n-grams of varying lengths as features – for instance, using both bigrams and trigrams simultaneously to capture character patterns at different scales. For function words, we use a manually curated list of dictionary forms and identify their occurrences using lemmatization data.[16] The POS tags are 9-character codes encoding grammatical information of words, such as part-of-speech, person, number, etc. (for the schema, see Celano 2017). Orthographic variation is encoded through paired tokens representing spelling variants (e.g.,

---

[15] For the sake of linking data, we also mark the possible TM Person ID (see https://www. trismegistos.org).

[16] For function words, see footnote 7. As mentioned, the line between function words and content words is not clear-cut, as e.g. adverbs may belong to either group (or some adverbs to function words and others to content words). In our study, the function word list contains definite articles, conjunctions, particles, adpositions and pronouns, as well as some adverbs (such as μάλα or πάνυ 'very').

'ου|ωι'), with multiple variations in a text represented as sequences, similar to the other features discussed above.

To convert documents into a format suitable for machine learning algorithms, we transform them into numerical feature vectors using Term Frequency-Inverse Document Frequency (TF-IDF) vectorisation. This method weights tokens based on their frequency within individual documents while accounting for their distribution across the entire corpus, also normalizing for document length differences.

For author profiling, we use K-means clustering to examine natural document groupings based on the TF-IDF features. The algorithm groups similar items together into a specified number (k) of clusters, trying to make items within each cluster as similar as possible to each other, while keeping different clusters clearly distinct from one another. We set k to match our known author count to evaluate cluster-author correspondence. To visualise these high-dimensional TF-IDF vectors, we project them into two and three dimensions using Truncated Singular Value Decomposition (TruncatedSVD).

In the attribution task, we use a Logistic Regression classifier with L2 regularization, using balanced class weights to account for authors having different numbers of texts in our dataset. The pipeline first transforms the features using TF-IDF vectorisation, followed by feature scaling to account for their different ranges. We then conduct three-fold cross-validation by dividing the dataset into thirds, with each portion alternating as test data while the remainder trains the model. Performance is evaluated through precision, recall, and F1-scores, averaged across all three splits.

For all of the machine learning implementations discussed above, we use Python's Scikit-learn library (Pedregosa et al. 2011); for the visualizations, we use Matplotlib (Hunter 2007).

4 Case studies and tests

As case studies, we have selected texts from three archives or archive groups that include several documents and many tokens written and/or authored by the same people. In some cases, we also have information about texts written in the same hand. The texts from two of the archives have been linguistically annotated (treebanked), so the POS tags and lemmatisation are curated. On the other hand, in one case study (Pathyris archives), treebanked texts are only partly available, and we rely mostly on n-grams, function words and spelling variation. We focus

on the authors in this study because in the current dataset there is only one writer who has produced a significant number of tokens and belongs to the same archives as the authors.[17]

4.1 Introducing the archives, their writers and authors

The Zenon archive is the biggest and most famous archive from Ptolemaic Egypt, with approximately 2,000 texts.[18] The documents were written in the third century BCE and consist of many different types of texts. For the PapyGreek corpus, we have treebanked a selection of letters and petitions and marked up metadata on authors. That information is based on what the texts themselves reveal; they are written by several different hands, but further information on the hands is not available for all texts. The two authors who provide us with the most tokens are Amyntas (ID 380), with 1,106 tokens, and Apollonios (ID 134), the finance minister and Zenon's employer, with 1,243 tokens. They both composed several documents, and it is clear that many of them were written down by several different professional secretaries, although Amyntas's autographs and some of his linguistic features (and variation within) can be discerned (see especially Evans 2010).

The Pathyris archives (2nd–1st centuries BCE, see Vandorpe & Waebens 2009) consist of notarial sale, loan, and renunciation contracts. They come from several different family archives, but here we have a selection of the notarial documents. They have mostly not been treebanked, but the author and writer metadata has been added according to the previous study of Vierros (2012). Several documents written by same hands have been recognised (writer ID's), but the writer identification is a very complicated case, since the handwritings are very similar to one another. The texts are also very similar, including formulaic language known and used by all the different notaries in the area. Some individual features could be identified, as mentioned above.[19] The authors chosen for this study are those from whom we have the most tokens: Hermias (ID 155) with

_____

[17] Apollonios (ID 105) from the Katochoi archive has 5,676 tokens, but the next best candidate only has 866 tokens.

[18] The archive bears the identification number 256 in the Trismegistos Archives portal, and a concise description of the archive with main bibliography is available in Vandorpe (2013). For the language of individuals in the Zenon archive, see, e.g., Evans (2010; 2012; 2020).

[19] See Vierros (2012: 90-9) on comparing the handwritings in different documents signed by the notaries and the formulaic differences; the rest of the book analyses different linguistic features and their (possible individual) variation

7,659 tokens, Ammonios (ID 211) with 2,527 tokens and Paniskos (ID 156) with 1,782 tokens. In addition, we included Areios (ID 206) who has been marked as an uncertain author for over 1,000 tokens, although as a certain author only for 937 tokens. Hermias was a successor of Ammonios in the Pathyris agoranomic office, and Paniskos was a notary in the office of Krokodilon polis, who also was in a supervisory role to the notaries in Pathyris.[20] We should also note that many documents in which Hermias was the author (the notary) were written by one identifiable hand (writer ID 214, the so-called 'Hermias-hand'). Some were written by a different hand, though, and for some documents we have no knowledge about the hand.

The Katochoi archive (the one we started this chapter with) has been studied from the linguistic point of view recently by Bentein (2015) and Vierros (2020). Bruno's chapter in this volume also includes material from this archive. It dates from the middle of the second century BCE and contains petitions, letters and other types of texts that have been treebanked for the PapyGreek corpus, with the exception of the receipts. The editor Ulrich Wilcken (1927: 115) identified the hands of Ptolemaios, his brother Apollonios and several scribal hands. As authors, but not as writers, we also have two girls, the 'twins' Taues and Taous, who always act together; in this study, they both appear under the ID of Taues (ID 107). Ptolemaios (ID 99), who was their protector, was also an author in many petitions (but not in all) that concerned the girls' matters, so some texts have all three as co-authors.[21] Apollonios (ID 105), the younger brother, often did not have the role of an author; on the contrary, he was the writer in several cases. However, he was not necessarily the writer who produced the final, 'official', document, since he drafted earlier versions of the petitions as well as copied the versions produced by professional scribes.

---

[20] See Vierros (2012: 82–9) on the notarial offices and their relationships.

[21] There are differences in the co-authored texts in the way the petitioners present themselves. On some occasions Ptolemaios is the main author and the girls are only mentioned in the narrative part of the petition (such are UPZ 1.24, 1.52, 1.53). In others, Ptolemaios is named together with the girls as petitioners, often only by using the term 'twins' and not their names (UPZ 1.22, 1.33, 1.34, 1.51), though occasionally their names are also written out (UPZ 1.32). When the girls act without Ptolemaios, their names are always mentioned as authors.

Table 1. Authors included in the study, with document and token counts.

PapyGreek ID (Trismegistos Person ID)	Name	Archive	Tokens as certain author (texts)	Tokens as uncertain author (texts)
380 (769)	Amyntas	Zenon	1,106 (12)	
134 (937)	Apollonios	Zenon	1,243 (20)	
155 (73)	Hermias	Pathyris	7,659 (47)	201 (2)
206 (47)	Areios	Pathyris	937 (7)	1,103 (4)
211 (55)	Ammonios	Pathyris	2,527 (19)	198 (3)
156 (70)	Paniskos	Pathyris	1,782 (10)	277 (2)
99 (12813)	Ptolemaios	Katochoi	4,193 (20)	112 (1)
107 (343197)	Taues	Katochoi	2,578 (20)	804 (3)

4.2 Results of the profiling (clustering) tests

Here we test the profiling task, i.e., how well the texts (of the authors listed above) cluster together based on different features given to the algorithm. In the figures that follow, the colours represent the authors (identified by their ID numbers in the legend). Each document (act of writing) is represented by a randomly assigned symbol, and the shape of the symbol indicates which cluster it has been assigned to by the algorithm. In other words, the triangles form one cluster, squares another cluster, and so on. If the markers share the same colour, they are authored by the same person (as pre-assigned in the PapyGreek metadata). Simply put, if a plot shows, e.g., all triangles in the same colour and this colour is not found with other symbols, then we have a perfect result. In such cases, all texts by the same author (as we know from our metadata) have been correctly grouped together by the algorithm (which does not know the author IDs).

In the first plot (Figure 1), all authors are included (excluding those uncertainly assigned) using only character n-grams (n=2–5) and texts containing at least 50 tokens. The plot shows clusters where we see the different archives separated from each other. The texts from the Zenon archive are quite tightly packed together. The Katochoi archive texts are slightly more dispersed but still close to each other. The Pathyris archive texts, on the other hand, are dispersed much more widely. In fact, the clusters include different types of texts, not necessarily texts by different authors.

In the second plot (Figure 2), we added function words to the previous setup, and the result looks somewhat different from Figure 1. Again, the different archives cluster separately, but the Pathyris archives are more tightly packed
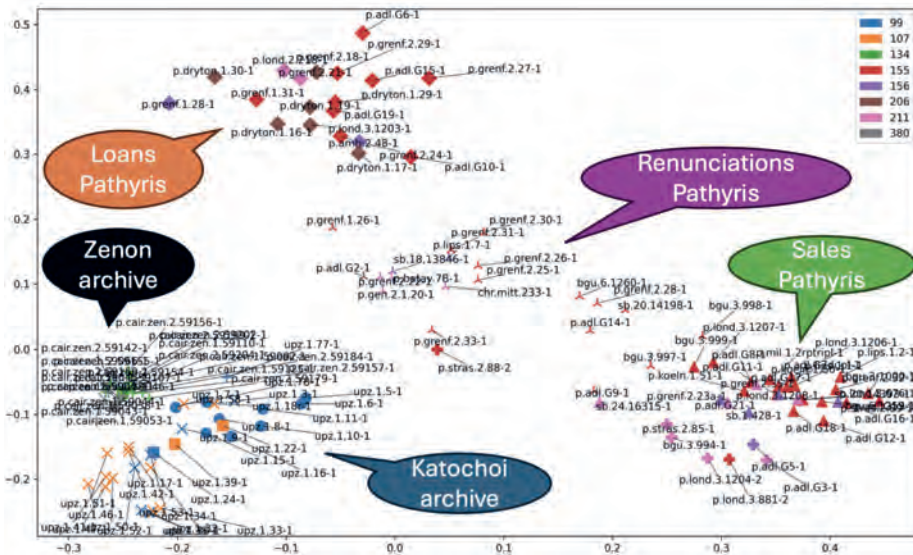
Figure 1. Text clustering results for all authors, based on character n-grams (n=2–5).

together on the left-hand side, with one set from those archives positioned further away. The author Hermias (ID 155) is present in both sets. In the case of the Katochoi and Zenon archives, they are in their own 'bubbles', each containing two different clusters that reflect the two authors, at least mostly (see below). The figure indicates that function word usage affects the clustering patterns.
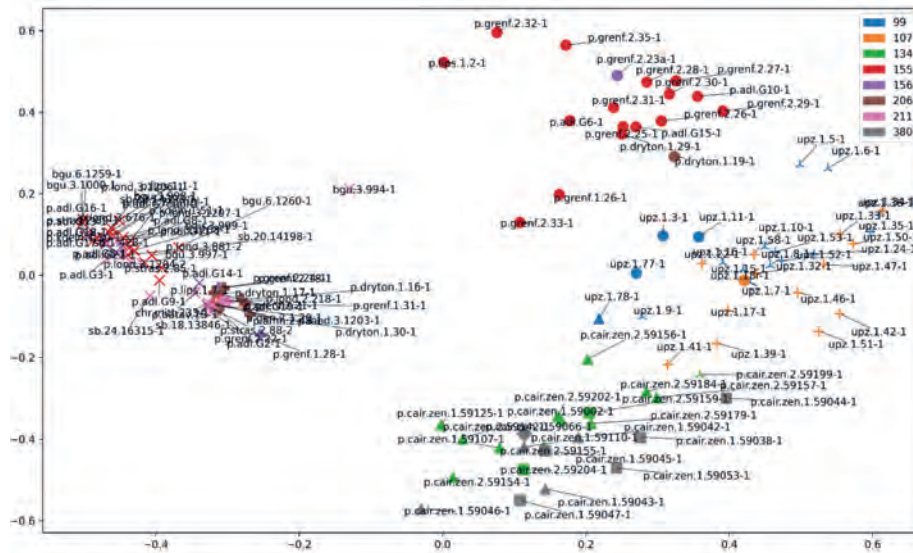

Figure 2. Text clustering results for all authors, based on character n-grams (n=2–5) + function words.

Since the archives cluster together quite clearly, it seems warranted to look at each individual archive or archive group separately; this could provide a more detailed picture of how authors of similar texts are clustered. First, concerning the notaries of Pathyris (including now also the uncertain authors), the character n-grams (n=2–5) produce a somewhat different plot (Figure 3) from when we combine function words and variation with the n-grams (Figure 4), although the colours (i.e., the named authors) are spread out in both: with character n-grams only, the clusters are clearly separated from each other, with different authors appearing in all of them randomly. This can be explained by the fact that the algorithm divides the clusters by document type: the circles represent loan contracts, the squares mostly represent sales contracts (with one outlier in the middle, a testament, P.Dryton 1.3). In the upper part of the plot are triangles; these include *homologia*-format contracts of slightly different contents: cessions and renunciations. It can be noted that the plot using n-grams (n=2–5) separates some documents in the sales cluster, whereas the bi- and trigram version (plot not presented here) separated two clusters in the homologia section (but not in the sales cluster).
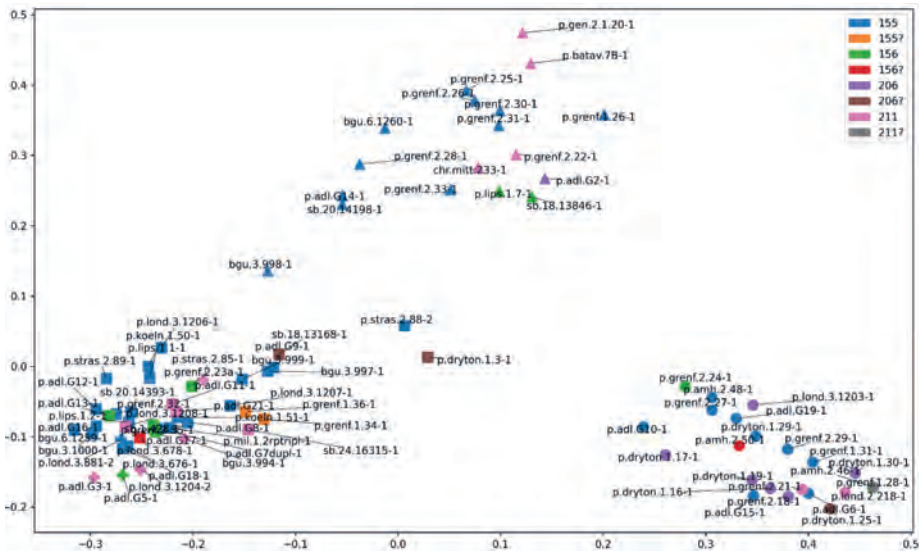

Figure 3. Pathyris authors, character n-grams (*n*=2–5.)

In Figure 4, the circles+crosses (bottom right) all seem to be loans, mostly authored by Hermias, whereas the triangle cluster (bottom left) also consists of loans, mostly authored by someone other than Hermias (though a couple of his
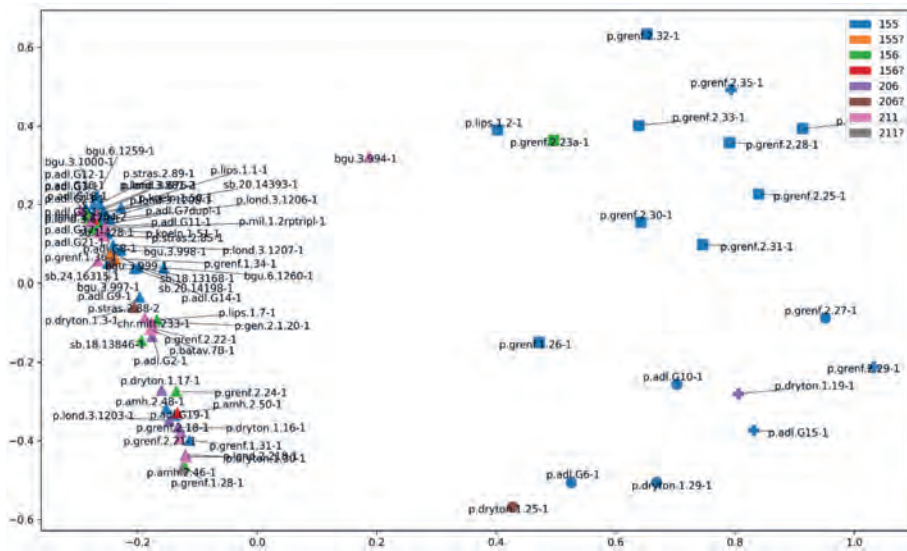
Figure 4. Pathyris authors, character n-grams (n=2–5) + function words + variation.

texts are there as well). One noteworthy detail concerns P.Dryton 1.25, which is uncertainly attributed to Areios. This text is very close to those of Hermias; Areios' name is not preserved in the text, nor is there information about the handwriting. Thus, for what it's worth, based on the algorithm the text would pass as one by Hermias. In the square cluster, there is some variation; some are cessions/renunciations of sales, others are sale documents (they may include similar phrases concerning renunciation). All except one are written by the notary Hermias. The exception, P.Grenf. 2.23a, is, unlike the others, written in Krokodilon polis by Paniskos, Hermias's superior. It is not clear why this document is placed within this cluster. Is there really something significant in the function word usage or orthographic variation, or is the decisive factor perhaps the protagonists of the contract (a group of four brothers with distinctive names who also often appear in other texts, usually written by Hermias and in Pathyris)? The cluster marked by a triangle mostly consists of sales documents, but it also includes some renunciations of sales. It must be concluded, therefore, that the clustering cannot draw exact lines between different document types that share common phraseology, such as sales and their renunciations. Loans are identified more clearly; however, when function word and variation analysis are added, the results differ, which is interesting considering the formulaic nature of the texts. In general, the texts by Hermias and by other notaries are, in fact, mostly clustered separately for some text types, so the author does play a role in the clustering as well.

For the Katochoi archive, several features and different numbers of n-grams were tested, and the results were in all cases quite good. Whether we used bi- and trigrams, a combination of n-grams (n=2–5) or 7-grams added with function words, variation and part-of-speech tags (POS tags) in isolation or in combination, the results differed from one another only in some details. We present only the best plot, based in this case on character 7-grams, function words, variations and POS tags (Figure 5). There are only two authors and, in addition to the texts certainly attributed to them, they both have some documents marked with uncertainty in their name. These two authors occasionally acted together, as mentioned above. The algorithm has identified two clusters, marked as triangles and circles. Most documents in both clusters are of the same types (petitions to the king, *enteuxeis*, or petitions to other officials, *hypomnemata*), so here the clusters are not divided by text type. The division follows the assigned authors almost perfectly: triangles representing Ptolemaios (ID 99, including the uncertain ones) and circles representing the twins (Taues ID 107, including the uncertain ones). Four twin-authored texts appear in the triangle cluster (see below). There are also two exceptional document types, dream descriptions, where Ptolemaios is the author (UPZ 1.77 and 1.78); these are both placed in the triangle cluster where other Ptolemaios's texts appear, but the one written by Apollonios, his brother, is positioned slightly further away. These two texts were closer together and further from the others in plots with different features (e.g., character bi- and trigrams, function words and variation, excluding POS tags).

The group of four documents that contains two of the 'wrongly' clustered texts actually consists of different versions of the same document,[22] and Ptolemaios and the twins are acting in it together. The drafts differ in how the petitioners are presented at the beginning. The twins' names are not specified, they are only referred to as 'twins' in UPZ 1.33 and 1.34. In UPZ 1.35 the mention of the twins is crossed out, leaving Ptolemaios as the sole petitioner, and in UPZ 1.36 Ptolemaios's name is the only one present. In the metadata, we have marked all three as authors of the first two. It is encouraging to see that they are very close to each other in the plot. The topmost text, UPZ 1.36, is the final official version of the *hypomnema* written by a professional scribe; the three remaining texts are written in the hand of Apollonios, Ptolemaios' brother, and are either drafts or copies. For example, UPZ 1.35 could be a copy, since its wording is very close to 1.36, but it shows more spelling variation. Texts 33 and 34 are of different format and shorter (unfinished), so they are more likely writing exercises that follow the original petition's wording but exhibit considerable spelling variation.

---

[22] Text group C in Vierros (2020: 50).

Two other texts (UPZ 1.22 and 1.32) are clustered in Ptolemaios's group but are marked as authored by Taues. These too are joint documents of Ptolemaios and the twins. UPZ 1.32 is the only text where all three are co-authors and where the twins' names are written out. Since Ptolemaios is mentioned first as author, the algorithm has recognised the text as belonging with others by Ptolemaios. UPZ 1.22 is similar to UPZ 1.33 and 1.34 discussed above, where the petitioners are listed as 'Ptolemaios and the twins'.

Two documents that are authored by Ptolemaios and in the triangle cluster (UPZ 1.52 and 1.53) appear to be very close in the plot to the circles, especially text 1.58, but when we look at the three-dimensional plot, we see that 1.52 and 1.53 form a distinct group. These two mention the twins in the text; in that sense, they are close to the circle cluster of the twins. They were also written by Apollonios, as were many of the petitions authored by the girls (without Ptolemaios). One element repeated in the circle cluster is, in fact, the names of the twins as senders of the petitions (exceptions being UPZ 1.51 and the ones mentioned above). It is noteworthy that the clustering algorithm finds similarities in the texts authored by the girls as opposed to the ones authored by Ptolemaios, even when the topics are similar and in both clusters Apollonios often appears as the writer (in Ptolemaios's cluster, also a bigger proportion of texts have been written by professional scribes). Apollonios was famous for his idiosyncratic morphology and spelling, so we can conclude that the writer is not the decisive
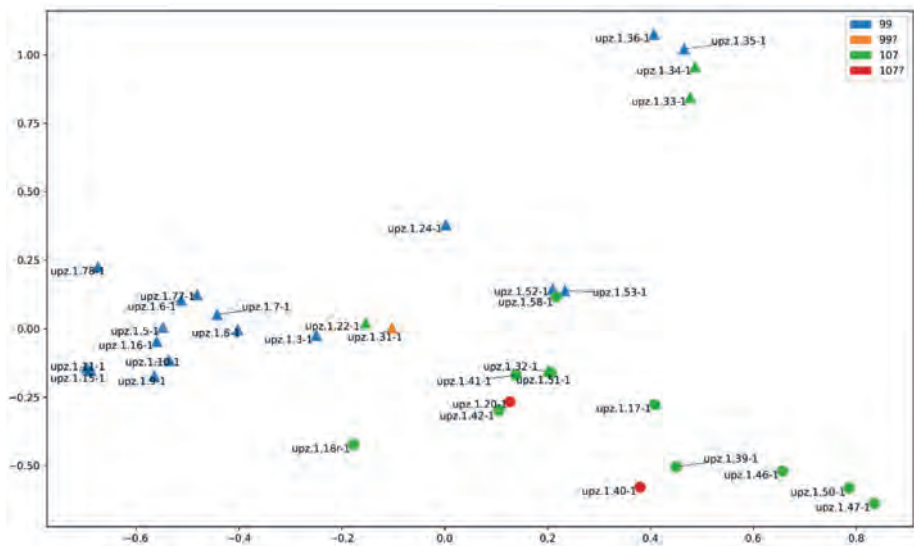

Figure 5. Two authors from the Katochoi archive, with character n-grams (n=7) + function words + spelling variation + POS tags.

factor for the algorithm, but the author is. At the moment, we cannot determine how much the names mentioned in the petitions, versus other textual features, influence the clustering results (this would require further testing and analysis of the model's decision-making processes).
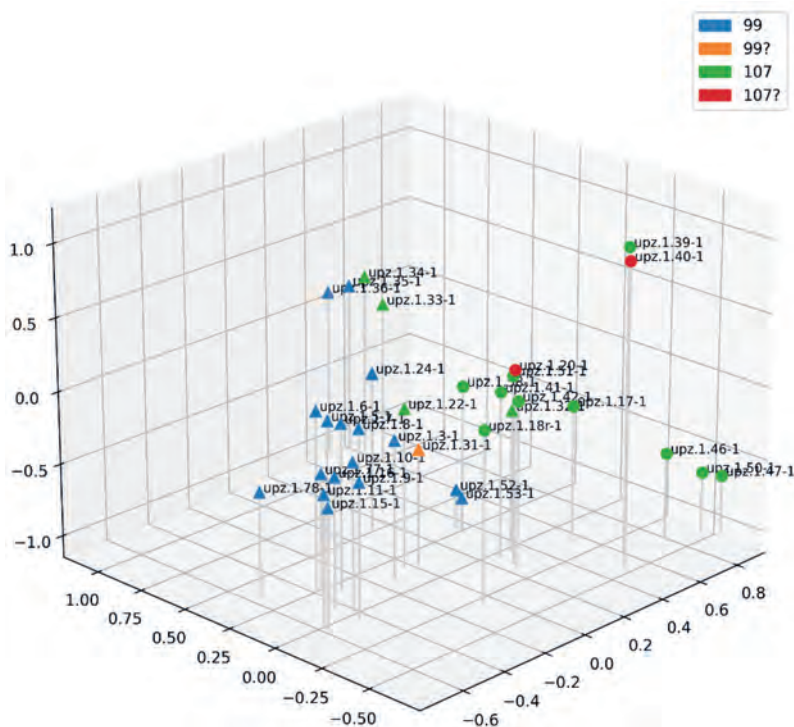

Figure 6. Two authors from the Katochoi archive. Same as Figure 5, but in 3D.

Moving on to the Zenon archive, there are texts from only two authors with no uncertain documents. Testing with character n-grams alone, the plot produced by n=7 shows a perfect division of clusters by author (Figure 7), with a clear division between Amyntas' (ID 380) orange circles and Apollonios' (ID 134) blue triangles. With n=2–5 we get some wrongly clustered texts, and adding function words does not improve the results; on the contrary, combining these with 7-grams produces a completely unsatisfactory clustering. Adding orthographic variation results in all texts being placed in a single cluster, except P.Cair.Zen. 1.50179 (both with 7-grams and n=2–5). Adding POS tags helps a little, but even taking only character n-grams and POS tags without function words or variation, the result is poorer than with plain character 7-grams.

Figure 8 is a 3D version of the same plot as Figure 7. Here we see even more clearly that three of Amyntas' texts are closer to those of Apollonios, while the rest form a group of their own at the right-hand side of the plots. All the letters in this test are quite short. When reading them through, certain issues emerge as possible explanations for the clustering.

The three texts closer to Apollonios' cluster, P.Cair.Zen. 1.59066, 1.59110 and 1.59053, are letters by Amyntas to Zenon. Normally, the letters by Amyntas employ more polite phrases, such as καλῶς ἂν (οὖν) ποιήσαις/ποιοῖς, 'please', as well as the conjunction ἵνα 'in order to' usually signalling a final clause. There is only one letter of Apollonios that has the phrase καλῶς (οὖν) ποιήσαις, P.Cair.Zen. 1.59179 (which, one may add, was placed very far from the others in the plots when the variation feature was included). It is also noteworthy that none of Apollonios's letters have a conditional clause beginning with the conjunction εἰ 'if' (it occurs only occasionally in Amyntas' texts, too). From the outliers in Amyntas's cluster, neither P.Cair.Zen. 1.59066 nor 1.59110 have the καλῶς ἂν (οὖν) ποιήσαις phrase or the conjunction εἰ. P.Cair.Zen. 1.59066 does have the conjunction ἵνα, which is mainly absent from Apollonios' letters (only present in P.Cair.Zen. 1.59142, which is close to Amyntas' letters in the plots). P.Cair.Zen. 1.59053, admittedly contains καλῶς ἂν (οὖν) ποιήσαις, but no conditional conjunction εἰ and twice the conjunction ἵνα. These features may only partially explain the clustering and placement in the plots from the
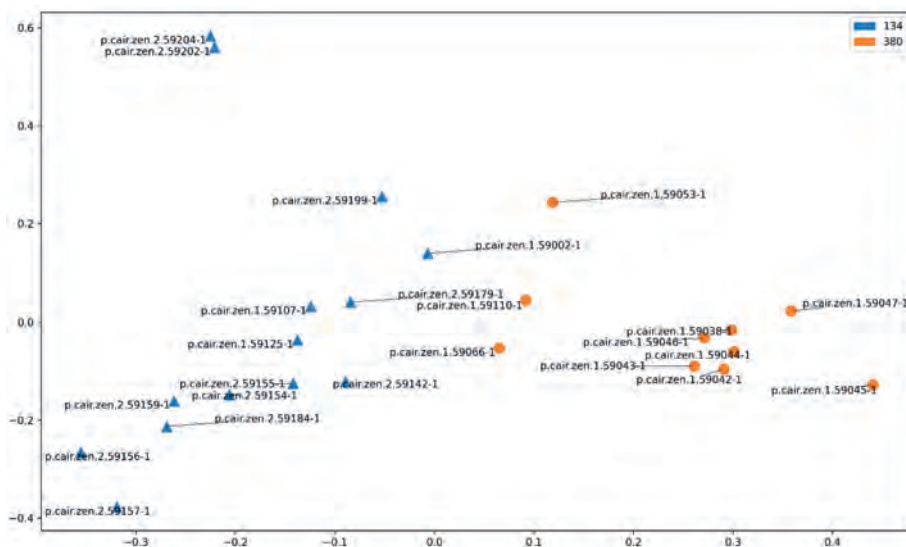

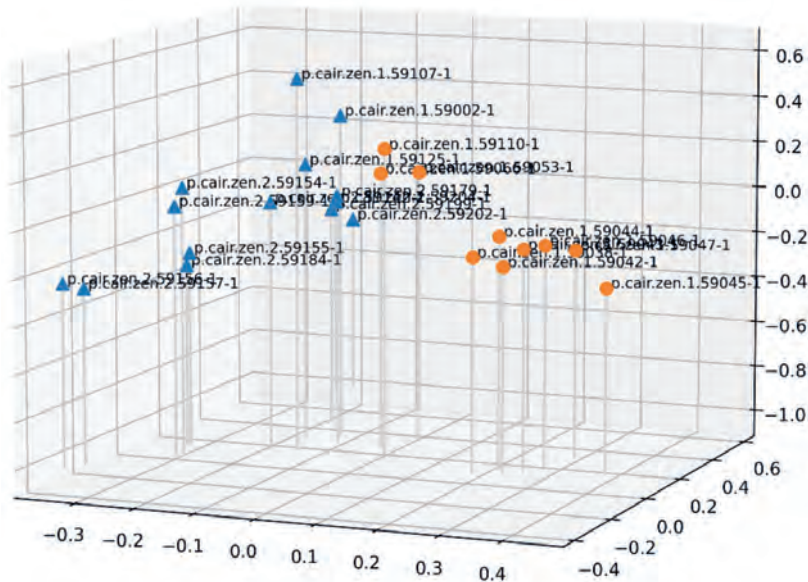Figure 7. Authors of the Zenon archive, with character n-grams (n=7).

Figure 8. Three-dimensional version of the same as Figure 7.

algorithm's perspective, but they appear to capture stylistic differences between the authors in these letters; the differences between these two authors can thus be explained by register variation, as Apollonios is a person of higher status than the other people in the Zenon archive and does not – perhaps due to his status – employ polite phrases.

4.2 Results of the attribution (classification) tests

In this Section, we turn to the text classification task: determining whether we can accurately predict authors of texts in our dataset. This is a closed-set classification problem, where the algorithm (discussed in Section 3.2) can only assign texts to one of the known authors in our training data. For this task, we experiment with different combinations of our features: n-grams, function words, variation and POS tags. To include function words, we need lemmatised texts, since our function word feature depends on lemmatised words (see again Section 3.2 for the details). Due to this limitation, this experiment includes only the following five authors with lemmatisations in the PapyGreek database: Taues (ID 107, 20 texts), Apollonios (ID 134, 20 texts), Hermias (ID 155, 9 texts), Amyntas (ID 380, 12 texts) and Ptolemaios (ID 99, 20 texts).

Table 2. Classification results using different feature sets

Features	Precision	Recall	**F1**
Variations	0.34	0.38	0.32
Function words	0.73	0.73	0.72
POS tags	0.81	0.79	0.80
N-grams (n=2)	0.90	0.88	0.88
N-grams (n=2–5)	**0.96**	**0.95**	**0.96**
N-grams (n=2–7)	0.95	0.94	0.95
N-grams (n=2) + function words	0.92	0.88	0.89
N-grams (n=2–5) + function words	0.96	0.95	0.96
N-grams (n=2–7) + function words	0.95	0.94	0.95
N-grams (n=2) + POS tags	0.92	0.92	0.92
N-grams (n=2–5) + POS tags	0.96	0.95	0.96
N-grams (n=2–7) + POS tags	0.95	0.94	0.95

Table 2 shows the classification results using three metrics: precision, recall and F1, macro-averaged across the possible classes (n=5). Precision indicates how many of the predicted attributions are correct; recall shows how many of the actual texts by each author were correctly identified. F1 combines precision and recall into a single score to show how well the model performs overall. All features are TF-IDF vectors, similar to the clustering experiments discussed in the previous Section.

The results indicate that the best feature type for this classification problem is the character n-gram, with the combinations of n=2–5 giving the best results (F1 score: 0.96). Notably, this score appears to represent a performance ceiling for our experiments, as adding function words or POS tags does not improve the result. However, when POS tags are combined with bigrams, they yield slightly better results than bigrams alone (F1 score: 0.92 versus 0.88). Using POS tags in isolation produces relatively good results (F1 score: 0.80), while function words alone perform more poorly (F1 score: 0.72). Finally, variations produce significantly lower scores (F1 score: 0.32), which is expected due to the limited data and the fact that most spelling errors and other editorially marked linguistic variants are not typically distinctive between authors, but rather follow general patterns in the development of post-Classical Greek in Egypt.

Figure 9 shows the confusion matrix of the predictions and true labels for the character n-gram feature set (n=2,3,4,5), in percentages. The confusion matrix displays the proportion of texts attributed to each predicted author (columns)

relative to their true author (rows). As the distinct diagonal pattern shows, the classifications are very accurate for all included authors.

For comparison, Figure 10 shows the confusion matrix for two other analyses using only function words (left panel) or POS tags (right panel). Hermias (155) is very accurately classified in both cases, which can be explained by two

Figure 9. Confusion matrix of author attribution results (values in %) using character n-grams (*n*=2–5).

factors: the very formulaic nature of his texts (e.g., many start with the date 'ἔτους ιδ/ιγ/ις…' etc.; see also Section 4.1) and the fact that he is the only one among these five authors who writes contracts. In both analyses, the Katochoi archive authors, Taues (107) and Ptolemaios (99), are well attributed, with Ptolemaios being especially distinguishable by function words. Amyntas (380), in contrast, is confused with other authors in both cases, and Apollonios (134) is more distinguishable by POS tags than by function word usage. This may reflect register variation mentioned above; for example, Apollonios seems to use more imperative forms.
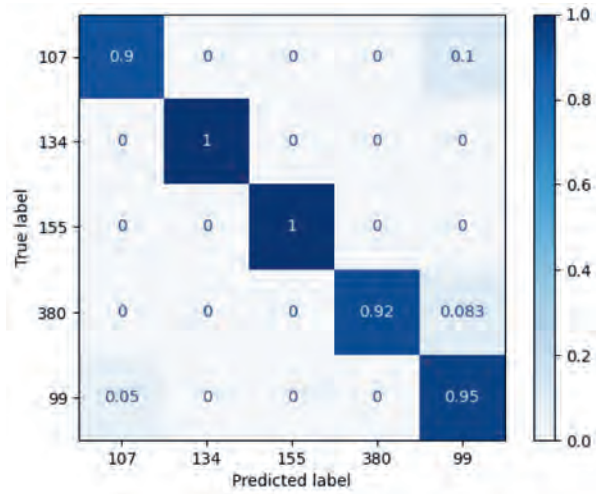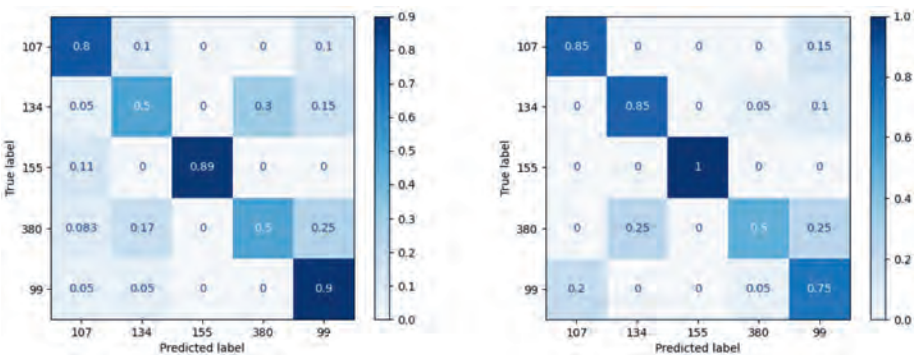

Figure 10. Confusion matrices based on function words (left) and POS tags (right).

5 Some conclusions and discussion

In this chapter, we experimented with authorship attribution algorithms on short and fragmentary Ancient Greek papyrus documents. The authors of these texts, as well as some of their writers, were largely known to us through textual and contextual information, allowing us to test whether computational methods could correctly identify authors. We aimed to investigate how profiling (clustering) and attribution (classification) algorithms perform with this type of material. Despite the challenging nature of papyrus texts, these methods worked well in many cases, though with varying degrees of success across different archives and text types.

In the profiling task, our three test archives formed quite clearly separate clusters, although the Pathyris archives were more widely dispersed. When analyzing all archives together, the texts clustered primarily by text type rather than by author. To better identify authorship patterns, we conducted separate analyses of each archive, since each contained somewhat different types of texts (with some overlap). These individual analyses yielded surprisingly clear author-based clustering in both the Katochoi and Zenon archives, though neither was perfect. In the Katochoi archive, some misclassifications occurred where authors frequently collaborated as co-authors of the same documents. The Zenon archive achieved perfect clustering when using character 7-grams as a feature, but other combinations of features and smaller n-grams were less successful.

The notarial documents from Pathyris were somewhat more difficult to cluster. Since notaries learned their craft through apprenticeship, documents of certain text types contained very similar phrases, leading to clustering by text type rather than by notary. The texts by Hermias, however, formed a more distinct cluster compared to other notaries' texts. While Hermias had the highest number of tokens, earlier qualitative analyses had also shown his use of distinctive linguistic features (Vierros 2012), which the algorithm may have detected.

Another important finding in relation to the clustering test was that the writer played a surprisingly minor role. Although several texts shared the same writer (mainly in the Katochoi and Pathyris archives), authorship emerged as a more significant factor than the actual writers with their influence in spelling variation.

The classification results were particularly encouraging, with an F1 score of 0.96 using character n-grams (n=2-5), suggesting that reliable author attribution is possible with fragmentary and short ancient documents under certain conditions. While POS tags alone performed well (80% F1 score), adding them or function words to the n-gram features did not improve upon the baseline,

indicating that character-level features may be sufficient for author identification in our corpus. However, these results must be interpreted cautiously. First, the small sample size (only five authors with 9-20 texts each) raises concerns about potential overfitting with our logistic regression model. Additionally, many texts in our sample contained explicit author identification, potentially influencing the results. Within the Pathyris archives especially, professional conventions and formulaic language may be more influential than individual style.

As for the question of which textual features worked best for the profiling task, results varied across different archives. In general, character n-grams alone gave promising results (with various values of n, most often 2–5 or 7). Adding function words and variation often enhanced the clustering, except in the case of the Zenon archive. The results were encouraging as these methods worked well with short texts without requiring resource-intensive linguistic annotation, though POS tags did occasionally improve the results. Lemmatisation and the use of function words are also recommended. We did not fully explore the treebanked data, as we did not include syntactic dependencies in our analysis. Future studies could investigate whether these syntactic features would yield even better results.

All in all, both the clustering of texts and their attribution to correct authors achieved promising results, particularly in the classification task with its high F1 scores. However, these results must be interpreted cautiously when applied to material where authors and writers are unknown. As our study showed, clustering might reflect text type rather than authorship, or several authors belonging to a sphere of a professional entity, like a group of notaries. The case studies in this article had several advantages: the texts often contained the authors' names, came from well-documented archives, and represented specific institutional contexts. For these reasons, our results may not generalise easily to other types of papyrological evidence.

References

Papyrus editions are referred according to the *Checklist of Editions of Greek, Latin, Demotic, and Coptic Papyri, Ostraca, and Tablets.* https://papyri.info/docs/checklist.

Baayen, Harald, Hans van Halteren and Fiona Tweedie 1996. 'Outside the Cave of Shadows: Using Syntactic Annotation to Enhance Authorship Attribution', *Literary and Linguistic Computing* 11(3): 121–32. https://doi.org/10.1093/llc/11.3.121.

Bentein, Klaas 2015. 'The Greek Documentary Papyri as a Linguistically Heterogeneous Corpus: The Case of the "Katochoi of the Sarapeion" Archive', *Classical World* 108(4): 461–84.

Celano, Giuseppe 2017. *Lemmatized Ancient Greek Texts (v1.2.5). https://github.com/gcelano/LemmatizedAncientGreekXML*

Eder, Maciej 2015. 'Does size matter? Authorship attribution, small samples, big problem', *Digital Scholarship in the Humanities* 30(2): 167–82. https://doi.org/10.1093/llc/fqt066.

Evans, Trevor V. 2005. 'Valedictory Ἔρρωσο in Zenon Archive Letters from Hierokles', *Zeitschrift Für Papyrologie Und Epigraphik* 153: 155–8.

Evans, Trevor V. 2010. 'Identifying the Language of the Individual in the Zenon Archive', in Trevor V. Evans and Dirk D. Obbink (eds.), *The Language of the Papyri*. Oxford: Oxford University Press, pp. 51–70.

Evans, Trevor V. 2012. 'Complaints of the Natives in a Greek Dress: The Zenon Archive and the Problem of Egyptian Interference', in Alex Mullen and Patrick James (eds.), *Multilingualism in the Graeco-Roman Worlds*, pp. 106–23. Cambridge: Cambridge University Press.

Evans, Trevor V. 2020. 'Not Overstrong in His Greek: Modern Interpretation of 'Egyptian' Greek Texts in the Zenon Archive', in Martti Leiwo, Marja Vierros, and Sonja Dahlgren (eds.), *Papers on Ancient Greek Linguistics*. Commentationes Humanarum Litterarum. Helsinki: Societas Scientiarum Fennica, pp. 43–62.

Gorman, Robert J. 2020. 'Author identification of short texts using dependency treebanks without vocabulary', *Digital Scholarship in the Humanities*, 35(4): 812–825. https://doi.org/10.1093/llc/fqz070.

Gorman, Vanessa B. and Robert J. Gorman 2016. 'Approaching Questions of Text Reuse in Ancient Greek Using Computational Syntactic Stylometry', *Open Linguistics*, 2(1). https://doi.org/10.1515/opli-2016-0026.

Grieve, Jack 2007. 'Quantitative Authorship Attribution: An Evaluation of Techniques', *Literary and Linguistic Computing* 22(3): 251–70. https://doi.org/10.1093/llc/fqm020.

Henriksson, Erik, and Marja Vierros (2025). 'PapyGreek Search: Exploring the Language of Greek Papyri.' In *Digital Papyrology*, edited by Nicola Reggiani, Vol. 3. De Gruyter.

Hunter, J. D. 2007. 'Matplotlib: A 2D Graphics Environment', *Computing in Science & Engineering* 9(3): 90–5.

Koppel, Moshe, Jonathan Schler, Schlomo Argamon and Yaron Winter 2012. 'The "Fundamental Problem" of Authorship Attribution', *English Studies*, 93: 284–291. https://doi.org/10.1080/0013838X.2012.668794.

Manousakis, Nikos and Efstathios Stamatatos 2023. 'Authorship Analysis and the Ending of Seven Against Thebes: Aeschylus' Antigone or Updating Adaptation?', *Classical World*, 116(3): 247–274. https://doi.org/10.1353/clw.2023.0007.

Mendenhall, T.C. 1887. 'The Characteristic Curves of Composition', *Science*, 9(214): 237–49.

Mosteller, Frederick and David L. Wallace 1963. 'Inference in an Authorship Problem. A Comparative Study of Discrimination Methods Applied to the Authorship of the Disputed Federalist Papers', *Journal of the American Statistical Association*, 58(302): 275–309. https://doi.org/10.1080/01621459.1963.10500849.

Pavlopoulos, John and Maria Konstantinidou 2023. 'Computational authorship analysis of the homeric poems', *International Journal of Digital Humanities*, 5(1): 45–64. https://doi.org/10.1007/s42803-022-00046-7.

Pedregosa, F. et al. 2011. 'Scikit-Learn: Machine Learning in Python'. *Journal of Machine Learning Research 12*: 2825-30.

Sanderson, Conrad and Simon Guenter 2006. 'Short Text Authorship Attribution via Sequence Kernels, Markov Chains and Author Unmasking: An Investigation', in *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, pp. 482–91. EMNLP '06. Sydney: Association for Computational Linguistics.

Schler, Jonathan and Moshe Koppel 2003. 'Exploiting Stylistic Idiosyncrasies for Authorship Attribution.' https://www.researchgate.net/publication/2950237_Exploiting_Stylistic_Idiosyncrasies_for_Authorship_Attribution.

Stamatatos, Efstathios 2009. 'A survey of modern authorship attribution methods', *Journal of the American Society for Information Science and Technology*, 60(3): 538–556. https://doi.org/10.1002/asi.21001.

Stamatatos, Efstathios 2016. 'Authorship Verification: A Review of Recent Advances', *Research in Computing Science* 123(1): 9–25. https://doi.org/10.13053/rcs-123-1-1.

Vandorpe, Katelijn 2013. 'Zenon Son of Agreophon.' Trismegistos Archives. https://www.trismegistos.org/arch/archives/pdf/256.pdf.   [Retrieved October 20, 2024]

Vandorpe, Katelijn and Sofie Waebens 2009. *Reconstructing Pathyris' Archives: A Multicultural Community in Hellenistic Egypt.* Collectanea Hellenistica 3. Brussel: Koninklijke Vlaamse Academie van België voor Wetenschappen en Kunsten.

Vierros, Marja 2012. *Bilingual notaries in Hellenistic Egypt: a study of Greek as a second language*. Collectanea Hellenistica 5. Brussel: Koninklijke Vlaamse Akademie van België voor Wetenschappen en Kunsten. http://hdl.handle.net/10138/319233.

Vierros, Marja 2019. 'Copying practices in Ptolemaic Egypt: A discussion based on agoranomic contracts from Pathyris', *Tyche: Beiträge zur alten Geschichte, Papyrologie und Epigraphik*, 33: 207–30.

Vierros, Marja 2020. 'Idiolect in focus: Two brothers in the Memphis Sarapieion (II BCE)', in Klaas Bentein and Mark Janse (eds.), *Varieties of Post-classical and Byzantine Greek*. Berlin: De Gruyter Mouton (Trends in Linguistics. Studies and Monographs, 331), pp. 39–74. https://doi.org/10.1515/9783110614404-003.

Vierros, Marja, and Erik Henriksson 2021. 'PapyGreek Treebanks: A Dataset of Linguistically Annotated Greek Documentary Papyri.' *Journal of Open Humanities Data* 7: 26. https://doi.org/10.5334/johd.55.

Wilcken, Ulrich (ed.) 1927. *Urkunden der Ptolemäerzeit (*Ältere Funde). Bd. I. *Papyri aus Unterägypten mit 2 Tafeln*. Berlin: de Gruyter.

4. Infinitives at Work. Competing Patterns in Early Ptolemaic Papyrus Letters[*]

CARLA BRUNO

1 Preliminary remarks

The reduction in the variety of finite and nonfinite subordinating strategies present in the classical system of complementation is one of most remarkable developments in the history of Greek syntax (Joseph 1987: 366; Horrocks 2007: 620). The process ended with the loss of nonfinite (i.e., infinitival and participial) complements in the later stages of the language, where finite embedded clauses became the rule (Joseph 1983: 37),[1] and it entailed a drastic simplification of the finite strategies at work (both in the classes of complementisers and the verb stems alternating in the embedded sentence).

The shift from the old to the new system was accompanied by a reassessment of the factors governing the distribution of the different patterns, giving rise to transitional diachronic types. Accordingly, in postclassical times, in addition to 'assertivity' (Crespo 1984) and 'factivity' (Cristofaro 1996; 2008)[2] other socio-pragmatic factors such as 'formality' (Bentein 2015a: 2017) gradually became more relevant in the complement choice, as nonfinite patterns are, for instance, 'mostly extended in higher social contexts' (Bentein 2017: 31).

---

[*] This research was first developed within the project *Multilingualism and Minority Languages in Ancient Europe*, which has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 649307 and then supported by the PRIN project *Non-finite verbal forms in ancient Indo-European languages: diachronic, synchronic and cross-linguistic perspectives* (Prot. 2022RSTTAZ), funded by MUR.

[1] The status of finite complementation in the Proto-Indo-European system is still unclear. According to Cuzzolin (2014) it is a late Proto-Indo-European development. Kiparsky (1995) claimed that the category of complementiser was lacking in Indo-European, which had only adjoined subordinated sentences. But cf. Lühr (2008: 156), who considers '*that*-clause containing a relational element … part of the basic Proto-Indo-European system'.

[2] Both 'assertivity' and 'factivity' concern the truth-value of the predication: the former in terms of the attitude of the speaker, the latter in terms of presupposition of the event involved. As argued by Anand and Hacquard (2014), the two aspects may not necessarily coincide.

Against the scenario of such long-term diachronic developments, this paper deals with the first hints of change in the early hellenistic stage, where an increased variation in finite sentential complement types is encountered. The data and examples discussed were taken from a small corpus of documentary private writings ranging from the third to the second century BCE, which comprises fifty-two papyrus letters selected by White (1986) for the early Ptolemaic period,[3] about a hundred documents from the Zenon archive (261–229 BCE) collected by Edgar (1931) for the university of Michigan, and about seventy texts (mostly petitions, but also letters and dreams) from the *katochoi* of the Sarapieion archive (164–152 BCE) edited by Wilcken (1927).[4]

The analysis takes into account the distribution of embedded infinitival clauses (Section 2) and their major finite competitors, i.e., the ὅτι, ἵνα and ὅπως clauses (Section 3), in contexts where they function as obligatory constituents of the main predicate. The survey points to some of the circumstances under which the emerging finite patterns spread outside their core functions at the expense of infinitival syntax, shedding light on some factors that were crucial in the renewal of the Greek subordinating system. Papyrogical documentary sources offer privileged perspectives on the ongoing change through the preservation of scribal mistakes and corrections, which testify to the grammars competing in the user's competence.

2 Infinitives at work

When looking at the infinitival complements present in the Ptolemaic period, nothing seems to announce the progressive decline which led to their loss in the later stages of the language. Infinitival clauses are still by far the most

---

[3] Despite the limited number of items included, this collection can be considered representative due to the editor's selection criteria. It was designed in order to include a wide diversity of epistolary types (i.e., letters of recommendation, familiar correspondence, petitions) and producers' profiles, who were sampled according to gender, social status, education and ethnicity (White 1986: 3). It thus results in a 'structured representative corpus for the purposes of linguistic analysis' (Porter and O'Donnell 2010: 294).

[4] The corpus does not include technical texts such as the accounts of the *katochoi* archive (cf. also Bentein 2015b: 464). It was also enlarged, when necessary, through targeted lexical research within the material aggregated by the Papyri.info and Trismegistos text databases. All the passages discussed are provided with an English translation, which – when not otherwise specified – was done by the author.

common strategy of complementation, and they occur in each of the syntactic configurations in which they had been typically hosted since the classical stages, exhibiting raising (as in 1) and control (as in 2–4) phenomena.

(1) PSI 4, 403, 2–7 (third century BCE)
τὴμ μὲν ἐπιστο- | λὴν ἣν ἀπέστειλας οὐ- | κ ἠδυνάμην ἀναγνῶι- | ναι διὰ τὸ ἐξηλεῖφθαι· | ἐδόκεις δέ μοι περὶ τοῦ | κλήρου γεγραφέναι.
'I could not read the letter you sent me because it is ruined; it seemed you had written me (lit. 'you seemed to have written me') about the piece of land.'

(2) UPZ 1, 64, 7–8 (156 BCE)
καὶ ὁ ἀδελφός σου ἀνθωμολογεῖτο μὴ ἠδικῆσθαι ὑπ' αὐ- | τοῦ, καὶ παρεκάλεσα αὐτὸν ἔρχεσθαι, περὶ ὧν ἂν βούληται.
'and your brother confessed that he had not been harmed (?) by him, and I encouraged him to come to me, about whatever he wanted.' (trans. White 1986: no. 39)

(3) PSI 5, 502, 24 (257 BCE)
ἠξιοῦμεν αὐτὸν συμπαραγενέσθαι· ὁ δ' ἔφη ἄσχολος εἶναι πρὸς τῆι τῶν ναυτῶν ἀποστολῆι.
'I asked him to assist us; but he said that he was busy in the dispatch of sailors.' (trans. White 1986: no. 18)

(4) UPZ 1, 62, 19–20 (160 BCE)
καὶ ἀπέλυσα | εἴπας αὐτῶι ὀρθρίτερον ἐλθεῖν
'and I sent him off, telling him to come early the next morning' (trans. White 1986: no. 38)

Infinitives cannot check the agreement with their subject. Accordingly, in raising and control configurations, the interpretation of their null subject depends on an argument position in the higher clause. The two patterns differ in that in raising environments the subject of the embedded predicate appears in the higher clause as an argument of the matrix predicate, as in (1), where ἐδόκεις does not impose any semantic restriction on its subject, which is instead thematically linked to γεγραφέναι. Further evidence of its raising from the embedded predication is the possible correlation with equivalent impersonal structures (as 5 below), which exclude any relationship of the accusative (i.e., the intended infinitival subject) with the main verb.

Such pairs of sentences (cf. 1 and 5) are therefore traditionally taken as the 'personal' and 'impersonal' alternatives. Conversely, in control configurations, the null subject of the infinitive establishes obligatory coreference with one of the arguments of the matrix predicate (e.g., the subject in 2, the direct object in 3 or the indirect object in 4) via an empty syntactic position.[5]

Similar cross-reference phenomena between the infinitival subject and the arguments of the main clause are not uncommon for nonfinite constructions, which interlinguistically tend to be more integrated into the main predication (Noonan 1985). However, Ancient Greek infinitives are not confined to similar syntactic environments and they can also exhibit distinct subjects in the accusative case. The pattern – the so-called *Accusativus cum Infinitivo* – is illustrated in (5) and (6), where με (in 5) and Ζήνωνα (in 6) are, respectively, the intended subject of καταρραθυμεῖν and ἀδεικεῖν. Because of contexts such as (5), where the pattern occurs with an impersonal predication, which does not allow direct object positions,[6] the accusative case cannot be assigned by the main verb.

> (5) P.Cair.Zen. 3, 59408, 4–12 (third century BCE)
> καλῶς   ποιήσεις, | περὶ   ὧν   σοι   Εἰρη- | ναῖος
> ἐνετείλατο, | δούς   μοι   τὸ   ἐφόδιον, | ὅπως   ἂν   ἀνα- |
> κομισθῶ πρὸς αὐτὸν | καὶ μὴ δοκῆι με | αὐτοῦ καταρρα- | θυμεῖν.
> 'Please, with regard to the matter Eirenaios instructs you about, give me a
> traveling allowance, in order that I come upcountry to him and do not appear
> to neglect him.' (trans. Bagnall and Cribiore 2006: 99)

---

[5] Control environments in ancient Greek have been discussed in terms of both syntax (as in Sevdali 2013) and lexico-semantics (Joseph 2002). See also Philippaki-Warburton and Catsimali (1999: 153), where the modern Greek *να* finite complements are taken as evidence that the 'syntactic and semantic correlates' of the control 'are separable'.

[6] The infinitive licensing of an independent subject as well as its accusative marking are long debated aspects of Greek syntax, especially for the theoretical implications for the notion of finiteness. Ancient Greek shows nonfinite forms that are able to license their own subject and are responsible for case marking. Structures such as (5), where the infinitival sentence completes an impersonal predication, have been taken as evidence against the Exceptional Case Marking (ECM) hypothesis, cf., e.g., Spyropoulos 2005 and Sevdali 2013). The accusative case is assumed to be the default case for infinitival complements by Philippaki-Warburton and Catsimali (1997: 583), who find that 'most often' infinitival clauses are object complements. In more traditional (nonformal) approaches to the matter, a diachronic relationship is found between control configurations and the *Accusativus cum Infinitivo*, as the emergence of overt accusative subjects is traced back to the reanalysis of accusatives (licensed by the main predicate) controlling the infinitival complement (Hettrich 1992).

(6) P.Mich. 1, 58, 16–22 (248 BCE)

κατέσ- | τησα δὲ καὶ τὸν Ἐτέ-| αρχον ἐπὶ Νικάνορα| περὶ τούτων, καὶ ἔ- | φη Ζήνωνα ἀδει- | κεῖν αὐτὸν καὶ οὐ- | κ αὐτόν.

'I also brought Etearchos before Nikanor about this matter, and Nikanor said that Zenon and not Etearchos is wronging the man.' (trans. Edgar 1931: 134)

Consequently, in ancient Greek, linear sequences involving an accusative plus an infinitive can relate to different syntactic configurations, depending on whether the accusative is assigned by the matrix predicate to its direct object (as in control settings such as 3), or by the embedded predicate to its overt subject (as in the *Accusativus cum Infinitivo*).[7]

On the other hand, infinitival accusative subjects are not restricted to environments involving disjoint reference: coreferential accusative subjects are also available. They are not very common, and are generally considered to be 'emphatic' compared to the corresponding null subject constructions.[8] Remarkably, overt infinitival subjects emerge under circumstances similar to overt nominative subjects in finite clauses in pro-drop languages, where its expression is not compulsory, i.e., when 'they are discourse prominent or when they are distinct from a previous subject' (Sevdali 2013: 21).[9] For instance, in (7), an excerpt from the letter in which a certain Aristeides complains of having been given the unwelcome responsibility of commissary of grain, the infinitive προβεβλῆσθαι displays an overt accusative subject (i.e., με at line 3) coreferent with the dative μοι (at line 2) in the matrix clause, which usually involves a control pattern (as in 8). Here, the expression of the infinitival subject is consistent with the emotional tone of the writing, where the repetition of the first-person singular pronoun (cf.

---

[7] Note that given the syntactic relationship between the accusative item and the main predicate only control settings allow its advancement to subject in passive contexts such as (i):

(i) P.Yale 1, 42, 25–9 (229 BCE)

καὶ | τ[ού]των χάριν παρακατεσχέ- | [θη]ν ὑπὸ τοῦ διοικητοῦ, μ[ή-] | ποτε ἀξιωθεὶς ἐ[μφ]ανίσηι τῶι | διοικητῆι μὴ δύνασθαι ἀχθῆναι.

'and on this account (or, their account) I have been detained by the dioiketes, lest having been asked he might make clear to the dioiketes for he (they?) cannot be held (for trial).' (trans. White 1986: no. 28)

[8] Cf. Luraghi (1999) for a survey of the circumstances under which coreferential accusative subjects emerge in classical Greek.

[9] Note that null accusative subjects can also be assumed for ancient Greek either with arbitrary reference or, given the appropriate context, with an independent reference (cf. Sevdali 2013 for a discussion on data from the classical stage).

μοι at line 2; με, μοι, μοι at line 3; με at line 4) conveys the sense of despair of the writer.

(7) P.Mich. 1, 23, 2–4 (257 BCE)
συμβέβηκέμ μοι ὑπὸ τῶν πο- | λιτῶν προβεβλῆσθαί με σίτου ἐγδοχέα οὔπω ὄντι μοι τῶν ἐτῶν οὐδὲ γινομένης μοι τῆς λειτουρ- | γίας ταύτης, ἀλλὰ διὰ φθονερίαν τινές [με προέβαλλον].
'I have had the misfortune to be proposed by the citizens as grain-buyer although I am not yet of the right age nor due for that burden, but [have been proposed by] certain persons out of jealousy.' (trans. Bagnall and Derow 2004: no. 88)

(8) P.Mich. 1, 83, 4–8 (third century BCE)
συν- | έβη γάρ μοι πρὸς τὸν λόγον | ἀσχοληθῆναι διὰ τὸ | Θράσωνα καταπλεῖν | εἰς Ἀλεξάνδρειαν.
'(Forgive me if I have not written to you for several days, for) I was obliged to busy myself over the account as Thrason was sailing down to Alexandria.' (trans. Edgar 1931: 161)

Apparently, only the language of dream reports regularly allows infinitival coreferential subjects in contexts not characterised by a special emphasis. This often occurs with the complements of οἴομαι, which typically introduces the dream content (cf. 9).

(9) UPZ 1, 77, 18–25 (161–58 BCE)
τὸ ἐνύπνιον, ὃ εἶδον Παχὼν | κ. οἴομαι ἀρειθμεῖν με | λέγων ὅτι Θῶυθ (ἔτους) κ | ἥως κ. | (ἔτους) κγ Παχὼν δ. ᾤμην | ἐν τῷ ὕπνῳ ἐπεικαλεῖν με τὸν | μέγιστον Ἄμμωνα ἔρχεσθαι ἀ[πὸ] | βορρᾶ μου τριτος ὢν, ἥως παραγ[ί] νηται.
'The dream that I (Ptolemaios) saw on Pachon 20. I seem to be counting (the days of the month) Thoth of year 20 until the 20th day. Year 23. Pachon 4. I (Ptolemaios) seemed in the dream to be calling upon the very great god Ammon, calling upon him to come to me from the north with two other (gods).' (trans. Rowlandson 1998: no. 80)

These dream reports are acknowledgedly low register texts, which include a large amount of orthographic and morphosyntactic variation (Bentein 2015b). They were personal notes 'meant for practical purposes rather than to be read by an

addressee' (Bentein 2015b: 480), therefore more open to linguistic innovations. Thus, cases like (9), where the *Accusativus cum Infinitivo* reinforces the classical control pattern, could reflect a vernacular strategy, by which the writer eases the tracking of the infinitival subject by avoiding cross-clausal references. In particular, since these notes were written by Ptolemaios, who is acknowledged to be the less skilled in writing of the two brothers *katochoi* at Memphis, the pattern may also be related to his poor education (cf. Vierros 2021, for an overview of some idiolectal tendencies in his writings).

The result is a totally uncanonical phrasing, since in ancient Greek, the *Accusativus cum Infinitivo* complements typically show a lesser degree of integration into the matrix clause. They can display not only an overt different subject (as in 5 and 6), but also a distinct event time as in (10), where the infinitive refers to a prior event.

> (10) P.Mich. 1, 82, 4–9 (third century BCE)
> σπεύδει \δὲ/ περὶ Αἰγυπτίου τινὸς | τῶν ἐκ τοῦ Ὀξυρυγχίτου | ὧι ὄνομα Πᾶρις, ὃν φάσκει ἐπι- | λελοχέναι Ἀξάπην τὸν βασι- | λικὸν γραμματέα εἰς τοὺς | μαχίμ[ου]ς
> 'He (i.e. Diokles) is interested in a certain Egyptian called Paris, belonging to the Oxyrhynchite nome, whom he says Axapes the royal scribe has enrolled in the native soldiery' (trans. Edgar 1931: 161)

Conversely, in control structures such as (11) and (12), the infinitival complements are bound to display the same subject and the same temporal coordinates as the main predicate.

> (11) P.Cair.Zen. 5, 59816, 7 (257 BCE)
> ἐπεὶ [οὖν] αὐτὸ[ς] | οὐ δεδύν[ημαι πα]ραγενέσθαι διὰ τὸ ἐνωχλῆσθαι
> 'Therefore, since I myself have been unable to come because of being sick' (trans. White 1986: no.19)

> (12) P.Col. 4, 66, 19–21 (256–55 BCE)
> δέομαι οὖν σου \εἴ σοι δοκεῖ/ συντάξαι αὐτοῖς ὅπως τὰ ὀφειλόμενα |κομίσωμαι καὶ τοῦ λοιποῦ εὐτάκτωσίν μοι ἵνα μὴ τῶι λιμῶι παρα- |πόλωμαι ὅτι οὐκ ἐπίσταμαι ἑλληνίζειν.
> 'Wherefore, I entreat you, if it seems acceptable to you, to instruct them I am to receive what is still lacking and that henceforth they follow orders lest I perish of hunger because I do not know how act the Hellene.' (trans. White 1986: no. 22)

The cooccurrence in the ancient Greek complementation system of overt accusative infinitival subjects (as in 5, 6, 7, 9 and 10) next to more canonical null subject embedded infinitives (as in 1, 2, 3, 4, 8, 11 and 12) is a typological singularity (Sevdali 2013).[10] These two patterns can also alternate in the very same context, as in (13), where συντιμήσεσθαι 'to estimate' and ἐκχωρήσειν 'to cede' are controlled by the subject of the main clause, whereas εἶναι licenses a distinct accusative subject (ὁμολογίαν 'agreement').

> (13) PSI 5, 502, 20–1 (257 BCE)
> μετὰ δ᾽ ἡμέρας δ καθίσαντες εἰς τὸ ἱερὸν οὐκ ἔφασαν οὔτε δικαίως οὔτ᾽ ἀδίκως | συντιμήσεσθαι, ἀλλ᾽ ἔφασαν ἐκχωρήσειν τοῦ σπόρου· ὁμολογίαν γὰρ εἶναι πρός σε αὐτοῖς ἐκ τοῦ γενήματος | ἀποδώσειν τὸ τρίτον.
> 'after four days, taking up residence in the temple (i.e., they went on strike), they said they did not want to agree to any valuation, be it fair or unfair, but preferred to renounce their right to the crop. For they alleged there was an agreement between you and them that they would pay one-third of the produce.' (trans. White 1986: no. 18)

According to Spyropoulos (2005), in ancient Greek the suspension of the obligatory subject coreference (and the consequent expression of an accusative subject) is bound to the temporal features of the embedded infinitive. In fact, the *Accusativus cum Infinitivo* is quite exceptional with predicates imposing the same time of the event on the embedded infinitive (such as δύναμαι in 11 or ἐπίσταμαι in 12), whereas it is common both in structures such as (13) where – as expected with verbs of saying – the embedded clause encodes an event with temporal properties independent of the main predication, and structures such as (14) below, where the infinitive is provided with temporal features that are predetermined by the matrix predicate, as 'the event time of the infinitival clause is future-oriented with respect to that of the matrix clause' (Spyropoulos 2007).[11]

---

[10] On the exceptional status of the classical infinitival complementation, see also Fykias and Katsikadeli (2015: 42), who emphasise 'the relatively limited span of time in the history of Greek' in which 'this noncanonical state of affairs … was in force'.

[11] As singled out by Spyropoulos (2005; 2007), from ancient to modern Greek, embedded complement clauses are sensitive to this feature, which determines the choice of the complementiser, as well as the triggering or suspension of control patterns. Three sets of complement clauses are hence distinguished: 'independent', 'dependent' and 'anaphoric' complements, respectively provided with 'full', 'fixed' and 'null' temporal properties.

This occurs regularly with the complements of directive predicates such as ἐντέλλομαι 'command' in (14), which typically refer to states of affairs that are 'prospected', rather than 'asserted' (Crespo 1984), as they are bound to the accomplishment of the request expressed by the predicate governing them. They are hence related to environments with a greater integration between embedding and embedded clause, both in the time and subject reference, and can also host control phenomena (as in 3 and 4).

> (14) P.Cair.Zen. 5, 59816, 2–3 (257 BCE)
> ἐνετέλλετο Ἀπολλώνιος μάλιστα μὲν αὐτὸν διελθεῖν πρὸς σέ, εἰ δὲ μή, ἀποστεῖλα[ί] | τινα παρ' ἐμοῦ ὃς ἀναγγελεῖ σοι τὰ παρ' αὐτοῦ.
> 'Apollonios ordered that, if at all possible, I myself should go over to you or, if not possible, to send one of my people to relay his instructions to you' (trans. White 1986: no. 19)

Accordingly, the classical – semantic – inner articulation between 'declarative' (i.e. 'referential') and 'prospective' (i.e. 'nonreferential') infinitives (cf., e.g., Jannaris 1897: 484) can be then seen as an interpretative correlate of the interaction of two formal parameters concerning the temporal properties of the infinitive and the designation of its subject.[12]

Thus, looking at the private papyri of the Ptolemaic period, infinitival complements still persist in the full range of their classical usages. However, these types display marked differences in their distribution, so that a rather complex picture emerges when considering their frequency.

On the syntactic level, null subject complements prevail over the *Accusativus cum Infinitivo*, which accounts for a minor proportion (22 per cent) of the infinitival complements encountered (430 tokens in total): there is then a slight decrease compared to the rates of this pattern in the classical period, where, according to Vassiliou (2012: 595), it accounts for 30 per cent of infinitive complements. In particular, control is, by far, the most common strategy by which the (null) subject of an infinitive can be tracked (particularly subject and indirect object control), whereas raising ('personal') types like (1), which were

---

[12] The two types are variously treated both in traditional grammars and current studies in linguistics. Humbert (1960: 182), e.g., refers to them as complements of 'judgement' and 'action'. In accordance with Kurzova (1968), Rijksbaron (2006: 97) opposes 'declarative' infinitives to 'dynamic' ones, discussing the difference in terms of the referential value of the form: unlike dynamic infinitives, which express a 'potential state of affairs', 'declarative' infinitives represent 'actual' states of affairs.

not unusual in the classical period (Jannaris 1897: 485; Kühner and Gerth 1904: 53), are actually very poorly represented and, when available, their impersonal counterpart (cf. 5) is preferred. This reflects a well-known typical tendency of postclassical compositions, where impersonal syntax is particularly pervasive (Jannaris 1897: 485).[13] In these environments, infinitival complements are quite regular, as is the *Accusativus cum Infinitivo*. More than one third of its occurrences (i.e., 35 per cent; thirty-three out of ninety-four tokens) occur in impersonal environments, where their persistence may have been also reinforced by the retreat of 'personal' syntax. The early decline of raising structures may then be a factor contributing to the relative resilience of the *Accusativus cum Infinitivo* within papyri.[14]

Furthermore, on the discourse level, the *Accusativus cum Infinitivo* reveals remarkable correlations with genres and text types of different levels of formality: on the one hand it regularly associates, in petitions, with the high formality of the *hupographe*, i.e. the formal decision about the request submitted (11 per cent of the total), which may be revealing about its higher level status; on the other hand, it occurs in the less formal dream reports, where its use with the verb οἴομαι (see 9 above) accounts for 9 per cent of the patterns encountered.

> (15) UPZ 1, 23, 2–8 (162 BCE)
> τοῦ προκειμένο[υ ὑ]πομνήματος ἐπιδεδομένου Σαραπί[ω]νι | τῶν διαδόχων καὶ ὑποδιοικητῇ παρὰ Πτολεμαίου | τοῦ προεστηκότος τῶν ἐν τῶι μεγάλω Σαραπιείωι διδυμῶν | περὶ τοῦ καθήκοντος αὐταῖς ἐκ τοῦ βασιλεικοῦ κατ᾽ ἐνιαυτὸν | ἐλαίου σησαμίνου καὶ κίκιος ἔχωντος ὑπογραφὴν· Μεννίδει. | ἐπισκεψάμενον ὅσα καθήκει ἀποδοῦναι, παρὰ δὲ σοῦ· τοῖς γραμ- | ματεῦσι. ἐπισκεψαμένους ἀνενεγκεῖν.
>
> 'The above report presented to Sarapion, *hypodioikêtês* with the rank of *diadochos*, from Ptolemaios who is patron to the twins in the great Serapieion, on the subject of the annual allowance of sesame and kiki oil due to them from the royal treasury, came with the instruction: "To Mennides. Look into the question of how much is due to them". You then countersigned it: "To the scribes. Look into the matter and report".' (trans. Thompson 1988: 224–5)

---

[13] On the lesser formality of the impersonal syntax in postclassical language, cf. Hult (1990), who pointed out the persistence of the infinitive with impersonal expressions in the New Testament.

[14] This is consistent with the statistical analyses of Vassiliou (2012: 565), according to which, from classical Greek to the Koine, the *Accusativus cum Infinitivo* exhibits an overall stability in impersonal environments, while it shows decreasing frequency rates elsewhere.

While the persistence of the *Accusativus cum Infinitivo* is expected in formal – as well as in formulaic – contexts (cf. also Bentein 2017 on later papyrological data), its clustering in the dream reports recorded by Ptolemaios is quite remarkable. In particular, if its occurrence in the *hupographe* may reflect a conservative attitude of the official language, in the 'vernacular dream reports (*vulgaren Traumberichten*)' (Mayser 1926: 356), the pattern occurs in innovative environments replacing the expected subject control structure.

Other innovative aspects of the syntax of the Ptolemaic infinitives concern their modest advancements in other functional domains, such as in the complements of 'factive' predicates as a result of the decline of the participle (Horrocks 2010: 92). The following passages in (16) and (17), for example, respectively show participial and infinitival complements alternating in environments governed by γιγνώσκω 'know':[15]

> (16) P.Mich. 1, 10, 11–12 (257 BCE)
> γί]νωσκε δὲ ὑπὸ | τῶν χειμώνων κατενεχθέντας εἰς Πάταρα
> 'Know that they were driven in to Patara by the storms' (trans. White 1986: no. 12)

> (17) UPZ 1, 68, 2–3 (152 BCE)
> γίνωσκέ με πεπορεῦσθαι εἰς Ἡρα- | κλέους πόλιν ὑπὲρ τῆς οἰκίας.
> 'Know that I have gone to Herakleopolis about the house.' (trans. White 1986: no. 41)

---

[15] Note also that the opposite circumstance can be encountered, as in (ii) and (iii), where participial complementation unexpectedly occurs in nonfactive contexts with the adjective ἱκανός 'able' and the verb ἀξιόω 'ask'. The latter occurs in the formula typically introducing the request for reparation within the Sarapieion petitions instead of the regular infinitive ἀναγκάσαι, as it is generally understood by editors. In general, except for the polite formula with καλῶς and ποιέω, participial complements are sparsely represented in the texts under scrutiny and their occurrences are considered to be poorly informative for the purposes of this research.

> (ii) P.Cair.Zen. 1, 59060, 11 (257 BCE)
> ἀλλὰ σὺ εἰκανὸς εἶ διοικῶν ἵνα ἀποσταλῆι ὡς ἀσφαλέστατα
> 'but you are well able to manage that it be sent with the greatest possible security.' (trans. White 1986: no. 15)

> (iii) UPZ 1, 32, 34–5 (162–1 BCE)
> ἀξιοῦμέν σε | [ἀ]να[γ]κασας αὐτοὺς | ἀπ[ο]δ[ο]ῦναι ἡμῖν
> 'We ask you to force them to pay us'

A degree of uncertainty between infinitival and participial complementation is on the other hand suggested by the comparison of (18) and (19). They are taken from two versions of the petition by which the *katochos* Ptolemaios begs the royals to exempt his brother from military service. The scribe instructed to write down the text has in particular a problem with the syntax of ἐάω (Wilcken 1927: 177), which regularly governs the infinitive in the fair copy (cf. γίνεσθαι at line 22), but the participle in its draft (cf. ὄντα ἀλειτούργητον at line 24 and περὶ ἐμὲ ὄντα at line 25, both of which are subsequently corrected).

(18) UPZ 1, 15, 18–23 (156 BCE)
νυνὶ δὲ ὁ προ[γεγραμμένος] | Ἀπολλώνιος εἰς τὴν ἐμ Μέμφ[ει] ση[μέα]ν | πρώτην ἐντέτακται, ὑπὸ δὲ τῶ[ν ὑ]πη- | ρετῶν περισπᾶται εἰς τὰς λε[ι]τουργίας | καὶ κουκ ἐᾶται περὶ ἐμὲ γίνεσθαι, οὗ χάριν, | βασιλεῦ, σε ἠξίωσα.
'Now, the above-mentioned Apollonios has been assigned to the first body of troops in Memphis, and he is compelled to the service by the attendants and is not allowed to stay by me, that's why, King, I asked you.'

(19) UPZ 1, 16, 22–5 (156 BCE)
διὸ ἀξιῶ, Ἥλιε βασιλεῦ, μὴ ⟦με⟧ ὑπεριδεῖν με | ἐγ κατοχῆι ὄντα, ἀλλ᾽, ἐάν σοι φαίνηται, <προστάξαι> | γράψαι τῶι Ποσειδωνίωι, ἐᾶσαι αὐτὸν {ὄν} ⟦τα⟧ | ἀλειτούργητον \ἵνα/ περὶ ἐμὲ ⟦ὄντα⟧ \ἦι/
'Therefore, I ask you, Sun King, not to neglect me as I am in *katoche*, but, if you please, to give the order to write to Poseidonios to exempt him from military service (lit. 'to let him being free from military service'), so that he can stand by me'

Another aspect of postclassical developments of the infinitive concerns the increasing frequency of its 'articular' use, which represents approximately 15 per cent of the infinitival sentences encountered.[16] This is a typical feature of the Hellenistic Koine (Horrocks 2010: 94), developed from a classical strategy of nominalization, in which by means of the preposed inflected article the verb predication is forced into the functional domain of a noun phrase that may also be governed by a preposition. From its beginnings, the pattern has been seen as early evidence of the 'weakening' of the subordinating function of bare infinitives, which are 'morphologically strengthened by the addition of an extra

---

[16] The widespread use of the infinitive in papyri is not exclusively bound to 'the development of the articular infinitive' (Mandilaras 1973: 309).

particle' (Joseph 1983: 49–50). Far from being a simple paradigmatic variant of bare infinitival complements (as in the passages in 20–22 below), due to the possible variation of the preposition, the articular infinitive was particularly exploited in adverbial subordination environments from which bare infinitives would be otherwise banned (as in 23–6 below).[17]

> (20) P.Col. 3, 6, 3–4 (257 BCE)
> ἐπεὶ δέ με Ὀλυμπιχ[ὸ]ς ἐκώλυσεν τοῦ μὴ ἰδεῖν σε ε[ἰ]σεκομίσθην πρὸ[ς τὸ] | παιδίον
> 'And when Olymphichos hindered me from seeing you, I gained entry to the boy' (trans. White 1986: no. 10)

> (21) P.Cair.Zen. 1, 59015, 6–9 (259–258 BCE)
> καλῶς ἂν οὖν ποιήσαις τὴμ πᾶσαν σπουδὴν | ποιησάμενος τοῦ συλληφθῆναι αὐτοὺς | ⟦ἵνα καὶ οιᾳ . . οι⟧ καὶ παραδοὺς Στράτωνι | τῶι κομίζοντί σοι τὸ ἐπιστόλιον.
> 'Therefore, you would do well, making due haste that they be recovered, to hand them over to Straton who carries this note to you.' (trans. White 1986: no. 7)

> (22) P.Tebt. 1, 26, 14–19 (114 BCE)
> προσέπεσεν \ἡμῖν/ ⟦μοι⟧ | \πε[ρ]ὶ τοῦ/ ⟦[π]ερι του⟧ τοὺς ἐκ τῆς κώμης | [β]ασιλικοὺς γεωργοὺς ἐγκαταλεί- | [πο]ντας τὴν ἐπικειμένην | ἀσχολίαν ἀνακ[ε]χωρηκέναι | ἐπὶ τὸ [ἐν Ν]αρμούθι ἱερὸν
> 'I learned that the crown tenants from the village, having left their prescribed occupations, had retired to the temple in Narmouthis' (trans. White 1986: no. 47)

> (23) P.Cair.Zen. 5, 59816, 6–7 (257 BCE)
> ἐπεὶ [οὖν] αὐτὸ[ς] | οὐ δεδύν[ημαι πα]ραγενέσθαι διὰ τὸ ἐνωχλῆσθαι
> 'Therefore, since myself have been able to come because of being sick' (trans. White 1986: no. 19)

---

[17] Besides the bare purpose infinitive, which was particularly popular in postclassical Greek (Horrocks 2010: 94; Joseph 2002: 15, n. 26), but in stock since Homer (Wakker 1988), the infinitive also occurs in temporal and result adverbial clauses (respectively introduced by the overt complementisers πρίν and ὥστε).

(24) PSI 5, 502, 30–1 (257 BCE)

φανερὸν δέ σοι ἔσται ἐκ τοῦ τὸν σῖτον συναχθήσεσθαι μηδεμιᾶς | οὔσης ἐν τῶι τόπωι χορηγίας

'It will be obvious to you from the grain that will be gathered that there is clearly no bounty in the place' (trans. White 1986: no. 18)

(25) P.Yale 1, 42, 9–11 (229 BCE)

τὸ πλέον ἀγωνιῶν | ἕνεκα τοῦ μηδ᾽ ἕως τοῦ νῦν ἀκηκοέναι | τὰ κατά σε, πρὸς τὸν θεὸν συνεχρώμην πολλάκις.

'being the more anxious because up to the present I have heard nothing about your affairs, I consulted the god many times.' (trans. White 1986: no. 28)

(26) UPZ 1, 59, 12–14 (168 BCE)

ἐπὶ δὲ τῶι μὴ παραγίνεσθαί σε [π]ά[ντ]ων | τῶν ἐκεῖ ἀπειλημμένων παραγεγο[νό]τω\ν/ | ἀηδίζομαι

'but when you did not come back when all of the others who had been in seclusion returned, I was unhappy' (trans. White 1986: no. 34)

The passages from (20) to (26) illustrate the pattern. In (20–22), it is an alternative to the more common bare infinitive complements: in (20) and (21) the genitive case is assigned by the main predicate (respectively κωλύω 'hinder' and σπουδὴν ποιεῖσθαι 'be eager'),[18] while in (22) it is governed by the preposition περί 'about'. In this respect, the comparison between (22) and (27), taken respectively from the draft and the clean copy of the same letter, may be suggestive of the higher prestige of the bare usage, which is preferred to the articular infinitive in the final version of the text.[19]

---

[18] For possible bare infinitival variants, e.g., of κωλύω complements, see (iv) below, taken from the *incipit* of a *memorandum* to Zenon.

(iv) P.Cair.Zen. 3, 59493, 1–2 (third century BCE)
ὑπόμνημα Ζήνωνι παρὰ Πεμενάσιος. | κωλύει ἡμᾶς ὁ θυρουρὸς ἔρχεσθαι πρὸς σέ
'*Memorandum* to Zenon from Pemenasis. The door-keeper hindered us from coming to you'

[19] On the greater formality of the Accusative and Infinitive pattern in later postclassical and Byzantine documents, see Bentein (2015a; 2017).

(27) P.Tebt. 4, 1099, 2–4 (114 BCE)

προσέπεσεν ἡμῖν τοὺς ἐκ τῆς κώμη[ς βασι]λικοὺς γεωργοὺς | ἀνακεχωρηκέναι
ἐπὶ τὸ ἐν Ναρμούθι ἱερόν.

'I learned that the crown tenants of the village had retired to the temple which
is in Narmouthis.' (trans. White 1986: no. 46)

It is almost as if the scribe hesitated here between the two variants: in the draft
(cf. 22) the preposition and the article introducing the Accusative and Infinitive
are first deleted, then added again (and finally rejected in the clean copy). Similar
uses of the articular infinitive instead of the expected bare infinitive complement
are not uncommon, especially with verbs of communication, as shown in (28)
with γράφω 'write' and in (29) with ἀπαγγέλλω 'report'.

(28) P.Dryton 1, 36, 2–7 (130 BCE)

ἐπεὶ πλειονάκις σοι γρά- | φω περὶ τοῦ διανδραγαθήσαντα | σαυτοῦ
ἐπιμέλεσθαι μέχρι τοῦ | τὰ πράγματα ἀποκαταστῆναι, | ἔτι καὶ νῦν καλῶς
ποιήσεις παρα- | καλῶν σαυτὸν καὶ τοὺς παρ' ἡμῶν.

'Since I wrote to you often about acting consistently in a brave manner so as
to take care of yourself until matters return to the normal, so also once again
please encourage yourself and our people.' (transl. White 1986: no. 43)

(29) UPZ 1, 59, 25–7 (168 BCE)

ἔτι δὲ καὶ Ὥρου τοῦ τὴν ἐπιστολὴν παρακεκο- | μικότος ἀπηγγελκότος ὑπὲρ
τοῦ ἀπολελύσθαι σε | ἐκ τῆς κατοχῆς παντελῶς ἀηδίζομαι.

'Moreover, now that Horos, who brought the letter, has reported about your
release from possession (by the god), I am altogether unhappy.' (trans. White
1986: no. 34)

Due to the article and the preposition before the infinitive, these structures
are syntactically and semantically more transparent than their bare variants:
clausal boundaries are here made explicit, and the alternating prepositions
(like the consequent article case variation) specify the value of subordinate.
Their increased frequency in Ptolemaic papyri is consistent with the well-
known general diachronic drift of postclassical language towards greater
grammatical analyticity (cf., e.g., Joseph 1987: 360). This could also explain
why the *Accusativus cum Infinitivo*, which accounts for two thirds of the
articular usages, is so common within these structures, where the presence of
the article avoids the syntactic ambiguity with object control structures such

as (3), where the accusative is governed by the main predicate.[20]

The adverbial use with prepositions, however, overwhelmingly prevails among the articular infinitives, as in the passages in (23–6), which feature causal expressions respectively governed by the prepositions διά, ἐκ, ἕνεκα and ἐπί. Only occasionally does the articular infinitive appear in similar contexts without the preposition, as in (30), where the cause is expressed through the dative inflection of the article.

> (30) UPZ 1, 19, 13–15 (163 BCE)
> καὶ ὧδε μὲν οὐκέτι τολμήσαντος ἀποβῆναι, εἰς δὲ τὸν Ἡρακλεοπολίτην |
> χωρισθέντος, [τ]ῶι δὲ μὴ ὑμᾶς εἶναι σὺν αὐτῶι ὑπὸ τῆς ἀθυμίας μετήλλαχεν
> | τὸν βίον.
> 'he didn't dare disembark here anymore and he went off to the Herakleopolite
> nome. Because of our not being with him, he died from hopelessness.' (trans.
> Rowlandson 1998: no. 79)

Articular infinitival complements in the direct cases are almost completely absent. The one exceptional case singled out is quoted in (31), where ἐγκόψαι is introduced by τό. This is actually the most 'classical' of the articular uses, which is generally assumed as the context from which the pattern spread (Robertson 1919: 1064). Its decrease in the documentary language is apparently inversely proportional to the increasing use in the other – oblique – functions.

> (31) P.Mich. 1, 56, 4–5 (251–248 BCE)
> οὗ ἕνεκα εἵλκυσαι | ὑπὲρ ὧν δύο μνα- | ιείων γέγονεν αἴτιον | τὸ Ἀρίστανδρον
> ἡμῖν ἐγκόψαι.
> 'The reason why you have been kept waiting about the two hundred-drachma
> pieces is that Aristandros interfered with us.' (trans. Edgar 1931: 129)

---

[20] The article shows case alternation in accordance with the matrix verb or the preposition governing it. In later stages, τοῦ also occurs where the main verb does not impose a genitive case, gradually assuming 'the character of an element closely associated with the infinitival expression' (Mandilaras 1973: 334). Cf., for instance, (v), taken from the New Testament, where the 'pleonastic' genitive occurs before the infinitive.

> (v) Act. 10: 25
> ὡς δὲ ἐγένετο τοῦ εἰσελθεῖν τὸν Πέτρον, συναντήσας αὐτῷ ὁ Κορνήλιος
> πεσὼν ἐπὶ τοὺς πόδας προσεκύνησεν.
> 'As Peter entered the house, Cornelius met him and fell at his feet in reverence.'
> (New International Version transl.)

What emerges is thus a picture in which, compared to the classical period, besides the decrease in the incidence of some 'classical' patterns, the infinitive also displays an increased range of subordination possibilities:[21] it can replace the participle in 'factive' environments and spread to more adverbial types through its prepositional articulated uses. Note that these developments may be a reflection of the decline of the participle, as the adverbial (prepositional) infinitive is a possible alternative to the 'circumstantial' participle (Jannaris 1897: 483). The passage in (32), where the author corrects the prepositional articular infinitive with the genitive absolute, displays the two strategies competing in the same context for the expression of the causal subordinate.

(32) UPZ 1, 3, 3–5 (164 BCE)
τοῦ ζ (ἔτους) Θῶθυ καταφυγούσης [τ]ινὸ[ς ὀνόματι Ἡρακλείας] εἰ τὸ προγεγραμένον ἱερὸν καὶ \ο[ὔ]σης [. . . . .] / λε[ιτ]ουρ[γ]ο[ύσης μοι ⟦καὶ διὰ τὸ μὴ⟧ | (?)] ⟦ ἔχει με ⟧ \μηθένα/ τέκν[ο]ν ποησαμένου μου [α] ὐ[τ]ὴν
'During the month of Thoth, in the seventh year, as a certain Herakleia took refuge in the above mentioned temple and was in my service, since I have no children, I adopted her'

3 Beyond the infinitives. Exploring competing patterns of complementation

The comparison with the possible alternating finite patterns gives a more comprehensive overview of the status of infinitival complementation in Ptolemaic times. The types most frequently encountered compete with embedded infinitives from the earliest stages of the language:[22] i.e., ὅτι and ὅπως (mostly plus the indicative of the verb form), respectively available for 'declarative' and 'prospective' complements, and ἵνα plus the subjunctive, which had been a finite alternative to the purposive infinitive since Homer. Admittedly, the retreat of the infinitive is basically bound to the gradual spread of these patterns outside their original core functions: ὅτι plus the indicative verb gradually replaced infinitives

---

[21] Similar increases in productivity are not uncommon in the gradual retreat of the Greek infinitive (Joseph 2002: 16, n. 26), as for the so-called 'temporal' or 'circumstantial' infinitive, which 'represents an extension of the articular infinitive' in medieval Greek (Joseph 1983: 60).

[22] See Crespo (1984), who relates the formal variety of the classical system to the functional opposition among the alternating types. Conversely, Lightfoot (1975) claims that the different patterns are basically synonymous. Cf. Fykias (2014) for a summary of the debate.

in factive complements after verbs of 'cognitive activity – saying, thinking, and the like' (Joseph 1983: 38), whereas ὅπως and ἵνα were available for nonfactive (mostly 'prospective') complements after 'expressions of will' (Joseph 1983: 38).

Besides final subordinates (mostly involving the subjunctive form of the verb), ὅπως complement clauses are also common in classical times: they express the 'purposed object (*erstrebte Objekt*)' (Kühner and Gerth 1904: 372) of verbs of 'striving' and mostly entail an indicative (future) verb form. Apparently, due to the postclassical generalization of the subjunctive type, the formal opposition between final and completive ὅπως is neutralised. Correspondingly, in the corpus investigated the subjunctive is the only mood selected by ὅπως. Conversely, ἵνα complements are definitely a postclassical achievement: they were classically confined to final subordinates and gradually extended the scope of their distribution – analogously to ὅπως – to complement structures.[23]

The passages in (33)–(40) illustrate the range of contexts covered by the different patterns within the papyri under scrutiny, where classical uses (as, e.g., 33 or 35) coexist with constructions less likely to occur in the classical period.

> (33) UPZ 1, 70, 14–17 (152–151 BCE)
> γίνωσ\κε/ ὅτι πιράσεται | ὁ δραπέ[τ]ης μὴ ἀφῖναι | ἡμᾶς ἐπὶ τῶν τόπων | ἶναι
> 'Know that the runaway will try to hinder us from staying in these parts' (trans. White 1986: no. 42)

> (34) P.Cair.Zen. 1, 59060, 4–6 (257 BCE)
> Πτολεμαίωι δὲ φαίνεται, ὅσα κατ' ἄ[νθρωπον], | ὅτι τῶν νῦν ἀλιφομένων, οἳ προειλήφασιν χρόνον πολύν, πολὺ κρείττων π̣ [. . . . . .] | καὶ σφόδρα ὀλίγου χρόνου πολὺ ὑπερέξει αὐτῶν
> 'but it seems to Ptolemy, so far as a man can tell, that Pyrrhos is much better than those presently being trained, who started training a long time before him, and that very soon he will be much beyond them' (trans. White 1986: no. 15)

> (35) P.Cair.Zen. 2, 59241, 6–7 (253 BCE)
> καὶ ὡς ἂν τάχιστα γένηται, ἀπόστειλον εἰς Μέμφιν πρὸς Ἀρτεμί- | δωρον, πειράθητι ὅπως ἐν ἡμ(έραις) ιε γένηται.

---

[23] On the general and historical relation between purposive (final) subordination and infinitival clauses, see Haspelmath (1989). See Horrocks (2010: 129), on the possible role of Latin contact in the roman period in accelerating the enlargement of the scope of final ἵνα under the model of the wider ranges of uses of *ut*.

'So, as soon as it (the wool) is got, send it to Memphis to Artemidoros, and make an effort that it (the mattress) is completed within fifteen days.' (trans. White 1986: no. 23)

(36) UPZ 1, 62, 11–14 (160 BCE)
καὶ ἀξιώσαντός με, ὅπως, ἐὰν ἐνέγκηι | τρίτομον, μεταλάβωσιν αὐτῶι οἱ πα- | ρ᾽ ἐμοῦ γραμματεῖς πάντας τοὺς | χρηματισμούς
'and asked me whether, if he brought the third section of a roll, my scribes would undertake the transcription of all his business transactions' (trans. White 1986: no. 38)

(37) P.Col. 4, 66, 19–21 (256–255 BCE)
δέομαι οὖν σου \εἴ σοι δοκεῖ/ συντάξαι αὐτοῖς ὅπως τὰ ὀφειλόμενα | κομίσωμαι καὶ τοῦ λοιποῦ εὐτάκτωσίν μοι ἵνα μὴ τῶι λιμῶι παρα- | πόλωμαι ὅτι οὐκ ἐπίσταμαι ἑλληνίζειν.
'Wherefore, I entreat you, if it seems acceptable to you, to instruct them I am to receive what is still lacking and that henceforth they follow orders lest I perish of hunger because I do not know how act the Hellene.' (trans. White 1986: no. 22)

(38) P.Cair.Zen. 2, 59251, 6–7 (254 BCE)
καὶ τῆς οἰκίας δὲ τῆς ἐμ Φιλαδελφείαι | ἐπιμελόμενος, ἵνα ὡς ἂμ παραγενώμεθα καταλάβωμεν αὐτὴν ἐστεγασμένην.
'moreover, concern yourself with the house in Philadelphia, in order that I find it roofed whenever I arrive.' (trans. White 1986: no. 24)

(39) P.Cair.Zen 1, 59016, 4–5 (259 BCE)
φρόντι- | σον δὲ ἵνα καὶ τὸν Νικάδαν ἀποστείλης εἰς Βηρυτὸν μετ᾽ ἀσφαλείας.
'Make certain too that you send Nikadas to Beirut safely.' (trans. White 1986: no. 5)

(40) UPZ 1, 20, 53–9 (163 BCE)
προνοηθῆι δ᾽ ἵνα καὶ τῶν πρὸς ταῖς πραγματείαις | προσκληθέντων τῶν προσοφειλόντων τὸ ἔλαιον | καὶ τὸ κῖκι, ὁμοίως δὲ καὶ Ψινταέους τοῦ ἐπιστάτου | καὶ Ἀμώσιος τοῦ παρ᾽ αὐτοῦ τῶν ὁμοίως τὴν ὄλυραν | παρειλκυκότων οἷς συνετάγη ἀποδοῦναι, καὶ οὗτοι | ἐπ[α]ναγκασθῶσιν ἀποδοῦναι

'care should to be taken that – as the officers mentioned above in this regard are still owing the olive and castor oil, and, similarly, also the *epistates* Psintaēs and Amosis, his agent, although they were ordered to pay, delayed (delivery of) the rice-wheat – they also would be forced to pay'

Both in (33) and (34) ὅτι introduces an argumental sentence: it is governed by a verb form of γιγνώσκω 'know' in (33), and by an impersonal structure with φαίνεται 'it appears' in (34). Its use agrees with the classical norm in (33) with γιγνώσκω 'know', which does not exclude ὅτι clauses involving a 'factive' reading, while it is less expected in (34) with φαίνομαι 'appear' a prime 'nonfactive' predication mostly involving an infinitival sentential subject.[24] The words of Hierokles, the manager of a gym in Alexandria in the third century BCE, are thus a prelude to later developments, since the modern language also allows ὅτι clauses with φαίνεται (and φανερό) in similar contexts (cf. also James 2010 on the diachrony of comparable structures with δῆλόν ἐστι and δηλοῦται).[25]

The passages selected in (35)–(40), then, exemplify ὅπως (cf. 35–7) and ἵνα (cf. 38–40) complement alternatives to an infinitival sentence. All of them occur in contexts that mostly refer to the accomplishment of the directive act expressed by the main predicate. In these environments, two aspects are diachronically relevant: the use of ὅπως (plus the subjunctive) complements with new classes of predicates and the emergence of ἵνα complements. Accordingly, while in contexts such as (35) with πειράομαι 'try' ὅπως had been available since the earliest stages of Greek (Joseph 2002: 14), structures such as (36) and (37) are postclassical developments, in which an infinitival control configuration – here excluded by the presence of distinct subjects – is still the unmarked option both with ἀξιόω 'request' (as in 3 above) and συντάσσω 'order'. Finally, (38)–(40) illustrate the spread of ἵνα complements: they replace ὅπως in (38) and (39) with φροντίζω 'take thought' and ἐπιμελέομαι 'take care',[26] and establish themselves as paradigmatic variants of

---

[24] For a discussion of the ὅτι spread into 'nonfactive' contexts see also Bentein (2017), who worked on later documentary texts ranging from the first to the eighth century CE. On the gradual semantic bleaching that ὅτι underwent in the postclassical period, cf. also Cristofaro (1996: 159).

[25] Note that in modern Greek in impersonal structures the descendants of φαίνεται can take both indicative clauses introduced by ότι and subjunctive complements containing να (Ingria 1981: 196–8).

[26] Kühner and Gerth (1904: 377) refer to the following passage from Aristophanes as an exceptional ἵνα complement usage in the classical period. Accordingly, postclassical ἵνα developments may be considered to evidence the spread of a colloquial feature that was latent in the previous period.

the infinitive complements in (40), where the ἵνα clause can be assumed to depend on an impersonal matrix predicate.[27] Given the diachronic stability of the infinitive governed by impersonal or passive expressions (cf. also Hult 1990: 165), cases like (40) may be illustrative of their early rivalry with the infinite in documentary papyri. This is particularly relevant for ἵνα complements, which even in the later centuries are typical of the most informal levels of usage (cf. Hult 1990: 225).[28] On the other hand, ἵνα complements are generally assumed as lower level variants even for the Ptolemaic period, especially when alternating with ὅπως sentences (cf. Mayser 1926: 247; Clarysse 2010: 43).[29] The competition between ἵνα and ὅπως complements is a long and complex process gradually leading up to the loss of ὅπως in modern Greek. However, in contexts such as (41), taken from the correspondence between Artemidoros, the *dioiketes*' personal physician, and Zenon, the *dioiketes*' secretary, they can occur alongside each other through coordination.

> (41) P.Cair.Zen. 5, 59816, 4–6 (257 BCE)
> συνέτασσεν οὖν ἀναγγέλλειν σοι ἵνα ξυλοκοπηθῆι πᾶσα καὶ **ποτισθῆι** | κ[αὶ
> μάλιστ]α μὲν ὅπως κατασπείρητε πᾶσαν αὐτήν

> (vi) Ar. *Ach.* 653–4
> καὶ τῆς νήσου μὲν ἐκείνης | οὐ φροντίζουσ᾽, ἀλλ᾽ ἵνα τοῦτον τὸν ποιητὴν ἀφέλωνται
> 'they do not care of the island, but to take away our poet' (trans. after Sommerstein 1973: 78)

[27] Although the passage is open to different interpretations (cf., e.g., Wilcken 1927, who translates as 'and (Dionysios) takes care (*und damit (Dionysios) dafür sorge*)', the impersonal construction is more consistent with the style of the *petitum*, where impersonal expressions prevail, as shown in (vii), which reports the two lines immediately preceding (40). See Mayser (1926: 122), on the 'passive usage (*passive Gebrauch*)' of the 'deponent (*Deponens*)' προνοοῦμαι.

> (vii) UPZ 1, 20, 51–2 (163 BCE)
> τοῦτο ἡμῖν ἀπὸ τῆς ἐντεύξεως ἐπιχωρηθῆι, | ὅπως μηθενὶ ἐξῆι κατὰ τοῦτο ἡμῖν ἐνποδίζειν
> 'it may be granted to us because of this entreaty, that nobody can (lit. 'it is possible to nobody') hinder us like that'

[28] On the argumental status of ἵνα clauses in the passages discussed, see in particular (38), where the subordinate is apposed to a nominal item (i.e., τῆς οἰκίας 'of the house'), whose genitive case is assigned by ἐπιμελέομαι 'take care', thus proving its immediate dependency on the main predicate (Faure 2014: 174).

[29] Moreover, lexical factors may also play a role in the distribution of the two patterns: unlike ἐπιμελέομαι 'take care' (cf. 38), the corresponding noun-verb collocation ἐπιμέλειαν ποιοῦμαι combines almost exclusively with the ὅπως complement (cf. the passages reported in 43a).

'Therefore, he instructed me to tell you that it was all to be cleared of brushwood and irrigated and that, if at all possible, you should sow the whole of it' (trans. White 1986: no. 19)

Taken all together, these structures account for only 16 per cent of the complements encountered: there are forty-one complements introduced by ὅτι, twenty-one by ὅπως and twenty-two by ἵνα. However, despite their lower frequency compared with infinitival strategies, some conclusions can be drawn from the clustered pattern of their occurrences. On the discourse level, it is, for instance, worth noting that almost half of the ὅτι clauses singled out occur in the dream reports from the Sarapieion: this may be quite instructive of their association with more popular levels of expression (cf. also Wilcken 1927: 346). At the syntactical level, more remarkably, a common circumstance recurs in the set of structures exemplified in (33)–(40): they are all excluded from syntactic configurations involving obligatory subject and time coreference with the matrix predication, like (11) and (12) discussed above. Admittedly, these are the environments in which the infinitive persisted longer as the only complementation pattern:[30] they show the maximum degree of integration between matrix and embedded clauses, inheriting from the matrix verb not only their subject, but also their temporal coordinates.

Finite complements are instead available in contexts with partial or null integration in the matrix event, where the embedded predicate is provided with distinct temporal reference. In particular, ὅπως and ἵνα mark predications that – due to their prospective value – consistently take place in a time subsequent to the matrix event, whereas ὅτι occurs in complements that may freely vary in their time reference. For instance, in (42) below, the reference is to a past event.

(42) PSI 5, 502, 11–12 (257 BCE)
ἐκομισάμην τὴν παρου σου ἐπιστολὴν τοῦ Παχὼνς ιδ παρὰ Ζωίλου, ἐν ἧι γράφεις | θαυμάζων ὅτι οὐθέν σοι ἀπέσταλκα περὶ τῆς συντιμήσεως καὶ τῆς συναγωγῆς τοῦ σπόρου.

---

[30] Cf. Jannaris (1897: 487), who adopts semantic taxonomies, and Joseph (1983: 53–4) and Horrocks (2010: 93) in terms of syntactic configurations involving subject identity. Note also that Markopoulos (2009), in his diachronic survey of periphrases for the future, points out that verbs such as (ἐ)θέλω since classical Greek can take finite complements instead of the infinitive 'when the subject of the Infinitive is different from that of the verb' (Markopoulos 2009: 38). The pattern is assumed to be a feature of the spoken language, which in Hellenistic-Roman Greek spread, especially in papyri, also in coreferential contexts (Markopoulos 2009: 76).

'I received your letter on Pachon 14 from Zoilos, in which you express astonishment that I have sent you no word about the valuation and the gathering of the crops.' (trans. White 1986: no. 18)

Accordingly, in (34), the spread of ὅτι into an unclassical 'nonfactive' context may have been triggered by the free time reference of the verb form, since φαίνεται 'it seems' does not impose temporal restrictions on its complements. Here, the future orientation of the embedded predication depends only on the verb tense alternation (i.e., the future indicative ὑπερέξει 'he will surpass'), and is not related to the complementiser form (as in ὅπως and ἵνα types).

Accordingly, the postclassical spread of complement finite structures appears to be a function of the temporal values of the embedded predication. Finite strategies are banned from contexts with null temporal features and where they are available, their distribution is ascribable to the formal expression of the distinction – latent in the infinitival system – between sentences with 'free' and 'fixed' time reference, respectively marked by ὅτι and ὅπως/ἵνα.[31]

In this regard, and in view of their overwhelming preference for unlike subjects, the finite complements encountered function as paradigmatic variants of the *Accusativus cum Infinitivo*, whose decline is a 'striking feature of the official Koine' (Horrocks 2010: 93). Unlike subjects are routine with ὅτι complements[32] and they represent the majority of the ὅπως and ἵνα clauses,[33] where disjoint

---

[31] Cf. Horrocks (2010: 93), who describes the spread of finite complements in terms of an 'internal simplification of the language', according to which the two infinitival basic uses were gradually replaced by distinct strategies. See also Crespo (1984), who observes that classical finite clauses show semantic restrictions that do not apply to nonfinite complements.

[32] As in the *Accusativus cum Infinitivo*, in these environments also coreferential subjects emerge in particularly emphatic contexts, such as (vii), where the adverb ἰδίαι 'personally' places special emphasis on the subject, which is also overtly expressed by the personal pronoun.

(viii) UPZ 1, 36, 14 (162–1 BCE)
νομίσαντα ὅτι σὺ ἰδίαι χρηματίζεις αὐταῖ[ς]
'considering that you can personally provide for them'

(ix) UPZ 1, 24, 20–1 (162 BCE)
νομίσαντα ταῖς διδύμαις | ἰδίαι σε ταῦτα διδόναι
'considering personally giving these things to the twins'

[33] Note that the higher frequency of coreferential subjects in ἵνα-complements is also due to their recurrence in the formulaic (closing) wishes such as (x), which represent two thirds (i.e., eight out of twelve) of the coreferential subjects encountered with ἵνα.

reference is often the effect of the backgrounding of the possible coreferential subject by passivization as shown in (43a–b) with the collocation ἐπιμέλειαν ποιοῦμαι 'take care'.

(43a) P.Cair.Zen. 1, 59015, 30–34 (259 BCE)
ἀξιοῦντες | τὴμ πᾶσαν ἐπιμέλειαν ποιήσασθαι | ὅπως ἂν συλληφθῶσιν καὶ παραδῶι αὐτοὺς | Στράτωνι
'Therefore we wrote requesting that all effort be made that they be recovered and that they be handed over to Straton' (trans. White 1986: no. 7)

(43b) P.Hib. 1, 41, 20–25 (261 BCE)
ἐπιμέλειαν δὲ | ποίησαι ὅπως καὶ τὸ ὑπάρ- | χον ἔλαιον δι' αὐτοῦ ἤδη | πραθῆι
'taking care in addition that the existing store of oil be sold at this time' (trans. White 1986: no. 2)

Under the same circumstances, possible references to the dative argument of the main clause (as is usual for the classes of control predicates typically involved in these patterns) are apparently avoided. This is illustrated by the comparison between the passages in (44) both featuring an ἵνα complement governed by ἀναγγέλλω 'report': the dative item introduced by the main verb shows coreference with the embedded (grammatical) subject in the active environment in (44a), but not in (44b), where it corresponds to the logical subject of the passive embedded structure.[34]

(44a) P.Cair.Zen. 5, 59816, 10–11 (257 BCE)
ἀνήγελκα δὲ καὶ Ζήνωνι καὶ Ἀρτεμιδώρωι [τῶ]ι ἐμ Μ[έμφει], | καθότι μοι Ἀπο[λλώ]νιος ἐνετείλατο, ἵνα χορηγῶσιν ὑμῖν χαλκόν, ὅσου ἂν χρείαν ἔχη[τε] | εἰς ταῦτα.

(x) UPZ 1, 59, 30–1 (168 BCE)
χαριεῖ δὲ καὶ τοῦ | σώματος ἐπιμελ[ό]μενος ἵν' ὑγιαίνηις.
'You will favour me by taking care of yourself in order to stay well.' (trans. White 1986: no. 34)

On the relevance of epistolary formulas in the study of variation and change within papyri letters, cf. Stolk and Nachtergaele (2016).

[34] The frequency of the passive may here rest on pragmatic reasons. See Bruno (2020), for a discussion of its role in the mitigation of directive expressions by concealing the sender and the recipient of the request.

'Moreover, I have also informed Zenon and Artemidoros, who is in Memphis, about this, just as Apollonios instructed me, in order that they supply you as much copper as you need for these things.' (trans. White 1986: no. 19)

(44b) P.Cair.Zen. 5, 59816, 4–6 (257 BCE)
συνέτασσεν οὖν ἀναγγέλλειν σοι ἵνα ξυλοκοπηθῆι πᾶσα καὶ ποτισθῆι | κ[αὶ μάλιστ]α μὲν ὅπως κατασπείρητε πᾶσαν αὐτήν
'Therefore, he instructed me to tell you that it was all to be cleared of brushwood and irrigated and that, if at all possible, you should sow the whole of it' (trans. White 1986: no. 19)

The renewal of the classical complementation system thus starts with the development of an array of finite alternatives to infinitival environments in which due to disjoint time and subject reference the nonfinite form is minimally integrated into the main predication. Unsurprisingly, the prime contexts to suffer from the pressure of the concurring finite strategies are those clashing with the standard tendencies in complementation, where 'normally, infinitives are environments of control, and finite clauses are environments of disjoint reference and emphasis' (Sevdali 2013). As an effect of the ongoing change, the Greek system of subordination then aligned itself with the standard trends in the complementation syntax.[35]

Again, the scribal inaccuracies can be revealing of the competition between 'old' and 'new' grammar, as in (45), where finite and infinitival patterns occur alongside each other. It is an excerpt from the first petition submitted to the royals by the twins Taues and Taous, where the *petitum* is structured so that an unexpected *Accusativus cum Infinitivo* (cf. τὴν ἀποκατάστασιν … γενηθῆναι, at line 60) follows a sequence of subjunctives governed by ὅπως (cf. ἐπιχωρηθῆι, ἐξῆι, at lines 47–52, partially quoted in vii, n. 27, and προνοηθῆι in 40 at line 53). Editors generally report the irregularity of the passage and suggest taking the expression as finite. The text is generally attributed to the 'fine' hand of a scribe, and it is presumably a copy of the final version the petition (Vierros 2021): the shift to nonfinite syntax may then be regarded as the search for a more refined style in accordance with the formal register of the petition.

---

[35] Cf. Horrocks (2010: 93), who treats the accusative plus infinitive as the infinitival construction first replaced by finite devices in official Koine, particularly focusing on the marked status of accusative subjects in the Greek system of complementation.

(45) UPZ 1, 20, 58–60 (163 BCE)

καὶ τῆς Νεφόριτος καὶ | τοῦ υἱοῦ αὐτῆς κατασταθέντων ὑπὲρ ὧν ἐγκαλοῦμεν
| αὐτοῖς καὶ τούτων τὴν ἀποκατάστασιν ἡμῖν γενηθῆναι

'and after Nephoris and her son were summoned for the facts which we charge
them with, we will also obtain the reparation of the wrongs suffered'

4 Concluding remarks

Despite its persistence in the full range of classical usages (Section 2), between the
third and second century BCE, something changed in the distribution of infinitival
complements. The change concerns, on the one hand, the decreasing frequency
of complementation patterns involving raising (as the 'personal' counterparts of
impersonal structures) and overt accusative subjects (i.e., the *Accusativus cum
Infinitivo*) (Section 2), and, on the other hand, the development of a variety
of innovative finite strategies with overt complementisers (Section 3). All these
developments attest the same general preference for complementation strategies
involving an overtly distinct argument structure domain. Raising structures,
which involve merging arguments, are, therefore, largely dispreferred to their
impersonal counterparts and, remarkably, the diachronically more stable control
structures in most informal registers reinforce the classical pattern with the overt
expression of a 'redundant' accusative subject. Under such circumstances, the
Accusativus cum Infinitivo, which is the environment where the infinitive licenses
its own subject, turns out to be the basic strategy by which the speaker avoids
the 'classical' cross-clausal tracking of its null subject. Nevertheless, this pattern
suffers from the competition with finite strategies, where clausal boundaries are
marked by overt complementisers. They were clauses introduced by the ὅτι, ἵνα
and ὅπως complementisers, which had been at work within the Greek system of
subordination since the earliest stages of the language and whose gradual spread
in the postclassical and then the medieval stage definitively reduced the scope
of application of the infinitive. The survey of their distribution in a selection
of early Ptolemaic papyrus documents shows the relevance of their degree of
integration in the matrix clause in their spread at the expense of infinitival syntax.
Finite patterns are banned from syntactic environments with obligatory time
and subject coreference in relation to the main clause and are, instead, almost
exclusively confined to disjoint reference contexts. The prime contexts in which
finite complement variants are available are therefore those environments in
which an *Accusativus cum Infinitivo* pattern is more likely to occur. The retreat of

the infinitive thus involves the reduction of the scope of the pattern whose syntax – in view of the low integration in the main clause – corresponds less to the nonfinite status of the form. Accordingly, the more the complement predication is integrated in the matrix clause, the less the infinitive is prone to be replaced by the finite strategy: its retreat admittedly begins from *verba declarativa* and *sentiendi*, which typically involve less integrated complements, and later affects the complement of prospective predicates (Jannaris 1897: 484).

The variation among the different emerging patterns also depends on this factor: ὅτι occurs in complements that are fully independent in the time and subject designation and ὅπως/ἵνα in completives with a fixed time reference (assigned by the governing predicate, which may also entail control phenomena). The finite patterns then alternate according to the null or partial integration of the embedded predication in the main clause, as in the cases below in presence of the same governing verb γράφω 'write':

(46a) P.Cair.Zen. 1, 59015, 2–4 (258 BCE)
ἀνήγγελέν μοι Κρότος γεγραφέναι σε αὐτῶι | ὅτι οἱ παῖδες οἱ ἀποδράντες μηνυτρίζοιντο | εἶναι παρὰ τῶι Κολλοχ\ούτωι/
'Krotos informed us that you had written to him that the runaway slaves are with Kollokhoutos' (trans. White 1986: no. 7)

(46b) P.Col. 3, 9, 6–7 (257 BCE)
καλῶς δ' ἔχειν ὑπέλαβον καί σοι γράψαι ὅπως ἐάν τινά σου χρεί[αν τὰ πράγμα-] | τα ἔχηι συναντιλάβηι φιλοτίμως καὶ ἡμῶν ἕνεκεν καὶ Μενέτου.
'I have thought advisable to write to you also in order that, if the matter requires your assistance, you may colloborate zealously on our account and that of Menetos.' (trans. White 1986: no. 13)

In a diachronic perspective, this alternation formally expresses the distinction, latent in the classical system, between different levels of integration of the complement in the main clause.

Another aspect of the ongoing change concerns the reduction of accusative subjects under the pressure of the competing finite patterns, which involved a simplification of the processes available for designating the infinitival subject. Null coreferential infinitival subjects are far more common than overt accusative subjects, which (as an effect of the competing alternating finite devices) rarely occur. Remarkably, accusative subjects persist with articular infinitives (with or without a preposition), where the presence of an overt clause boundary marker

(i.e., the article) may have favoured the correct association of the accusative with the infinitival predicate domain.

The *Accusativus cum Infinitivo*, whose grammar clashes in many respects with the general tendencies of nonfinite syntax, is then found at the crossroads of diachronic drifts favouring constructional analyticity in the early postclassical Greek complementation system. One concerns the tendency towards the overt expression of clausal boundaries by means of the specialization of a set of particles that vary according to the degree of integration of the embedded predication, and the other the unification of the processes of subject designation in infinitival complements, where control structures survive by the infinitives licensing a distinct accusative subject.

Intricate, complex evolutionary paths are thus suggestively captured by the scribal practice, which by documenting variations as well as regularities in language, provides a unique insight for the understanding of language change.

References

Anand, Pranav and Hacquard, Valentine 2014. 'Factivity, Belief and Discourse', in Luka Crnič and Uli Sauerland (eds.), *The Art and Craft of Semantics: A Festschrift for Irene Heim*, 2 vols. Cambridge MA: MIT Working Papers in Linguistics, vol. I, pp. 69–90.

Bagnall, Roger S. and Derow, Peter 2004. *The Hellenistic Period: Historical Sources in Translation.* Malden, US: Blackwell Publishing.

Bagnall, Roger S. and Cribiore, Raffaella 2006. *Women's letters from Ancient Egypt, 300 BC–AD 800.* Ann Arbor: The University of Michigan Press.

Bentein, Klaas 2015a. 'Minor Complementation Patterns in Post-classical Greek (I–VI AD): A Socio-historical Analysis of a Corpus of Documentary Papyri', *Symbolae Osloenses*, 89(1): 104–47.

Bentein, Klaas 2015b. 'The Greek Documentary Papyri as a Linguistically Heterogeneous Corpus: The Case of the "*Katochoi* of the Sarapeion" Archive', *Classical World*, 108(4): 461–84.

Bentein, Klaas 2017. 'Finite vs. Non-finite Complementation in Post-classical and Early Byzantine Greek. Towards a Pragmatic Restructuring of the Complementation System?', *Journal of Greek Linguistics* 17: 1–34.

Bruno, Carla 2020. 'Forms of the Directive Speech Act: Evidence from Early Ptolemaic Papyri', in Ilja Seržant and Dariya Rafiyenko (eds.), *Postclassical Greek: Intersections of Linguistics and Philology*. Berlin – New York – Amsterdam: Mouton de Gruyter, pp. 221–44.

Clarysse, Willy 2010. 'Linguistic Diversity in the Archive of the Engineers Kleon and Theodoros', in Trevor V. Evans and Dirk D. Obbink (eds.), pp. 35–50.

Crespo, Emilio 1984. 'On the System of Substantive Clauses in Ancient Greek. A Functional Approach', *Glotta* 62: 1–16.

Cristofaro, Sonia 1996. *Aspetti sintattici e semantici delle frasi completive in greco antico*. Firenze: Nuova Italia.

Cristofaro, Sonia 2008. 'A Constructionist Approach to Complementation: Evidence from Ancient Greek', *Linguistics* 46: 571–606.

Cuzzolin, Pierluigi 2014. 'Subordination', in Georgios K. Giannakis (ed.), vol. III, pp. 335–41.

Edgar, Campbell C. 1931 *Zenon papyri in the University of Michigan collection*. Ann Arbor: University of Michigan Press.

Evans, Trevor V. and Obbink, Dirk D. (eds.) 2010. *The Language of the Papyri*. Oxford University Press.

Faure, Richard 2014. 'Argumental clause', in Georgios K. Giannakis (ed.), vol. I, pp. 172–8.

Fykias, Ioannis 2014 'Infinitives (Syntax)' in Georgios K. Giannakis (ed.), vol. II, pp. 229–36.

Fykias, Ioannis and Katsikadeli, Christina 2015. 'The Rise of "Subordination Features" in the History of Greek and their Decline', in Leonid Kulikov and Nikolaos Lavidas (eds.), *Proto-Indo-European Syntax and its Development*. Amsterdam – Philadelphia: John Benjamins Publishing Company, pp. 29–49.

Giannakis, Georgis K. (ed.) 2014. *Encyclopedia of Ancient Greek Language and Linguistics*, 3 vols. Leiden – Boston: Brill.

Haspelmath, Martin 1989. 'From Purposive to Infinitive. A Universal Path of Grammaticization', *Folia Linguistica Historica*, 10: 287–310.

Hettrich, Heinrich 1992. 'Die Entstehung des lateinischen und griechischen AcI. Rekonstruktion und relative Chronologie', in Robert Beekes, Alexander Lubotsky and Jos Weitenberg (eds.), *Rekonstruktion und relative Chronologie*. Innsbruck: Institut für Sprachwissenschaft, pp. 221–34.

Horrocks, Geoffrey C. 2007. 'Syntax: From Classical Greek to the Koine', in Anastasios F. Christidis (ed.), *A History of Ancient Greek*. Cambrige: Cambrige University Press, pp. 618–31.

Horrocks, Geoffrey C. 2010. *Greek: A History of the Language and its Speakers*, Second Edition. Oxford: Wiley-Blackwell.

Hult, Karin 1990. *Syntactic Variation in Greek of the 5th Century A.D.* Göteborg: Acta universitatis Gothoburgensis.

Humbert, Jean 1960. *Syntaxe grecque*. Troisième Édition. Paris: Klincksiëck.

Ingria, Robert Joseph Peter 1981. *Sentential Complementation in Modern Greek*. Unpublished Ph.D. dissertation, MIT University.

James, Patrick 2010. 'Variation in Complementation to Impersonal *verba declarandi* in Greek Papyri from the Roman and Byzantine Periods', in Trevor V. Evans and Dirk D. Obbink (eds.), pp. 140–55.

Jannaris, Antonius N. 1897. *An Historical Greek Grammar Chiefly of the Attic Dialect*. Hildesheim: Olms.

Joseph, Brian D. 1983. *The Synchrony and Diachrony of the Balkan Infinitive: A study in areal, general, and historical linguistics.* Cambridge University Press.

Joseph, Brian D. 1987. 'Greek', in Bernard Comrie (ed.), *The World's Major Languages.* London: Croom Helm, pp. 347–72.

Joseph, Brian D. 2002. 'On Some Control Structures in Hellenistic Greek: A Comparison with Classical and Modern Greek', *Linguistic Discovery*, 1: 1–16.

Kühner, Raphael and Gerth, Bernhard 1904. *Ausführliche Grammatik der griechischen Sprache*, 2 vols. Hannover – Leipzig: Hahnsche Buchhandlung, vol. II.2: *Satzlehre.*

Kurzová, Helena 1968. *Zur syntaktischen Struktur des griechischen: Infinitiv und Nebensatz.* Amsterdam: Hakkert; Prag: Academia.

Kiparsky, Paul 1995. 'Indo-European Origins of Germanic Syntax', in Adrian Battye and Ian Roberts (eds.), *Clause Structure and Language Change.* Oxford University Press, pp. 140–69.

Lightfoot, David 1975. *Natural Logic and the Greek Moods: The Nature of the Subjunctive and Optative in Classical Greek.* The Hague – Paris: Mouton.

Lühr, Rosemarie 2008. 'Competitive Indo-European syntax', in Gisella Ferraresi and Maria Goldbach (eds.), *Principles of Syntactic Reconstruction.* Amsterdam – Philadelphia: J. Benjamins, pp. 121–59.

Luraghi, Silvia 1999. 'The Subject of Complement Clauses with the Infinitive', in Bernard Jacquinod (ed.), *Les Complétives en Grec Ancien. Actes du colloque international de Saint-Etienne (3-5 septembre 1998).* Publications de l'Université de Saint-Etienne, pp. 199–213.

Mandilaras, Basil G. 1973. *The Verb in the Greek Non-Literary Papyri.* Athens: Hellenic Ministry of Culture and Sciences.

Markopoulos, Theodore 2009. *The Future in Greek: From Ancient to Medieval.* Oxford University Press.

Mayser, Edwin 1926. *Grammatik der griechischen Papyri aus der Ptolemäerzeit*, 3 vols. Berlin: de Gruyter, vol. II.1: *Satzlehre.*

Noonan, Michael 1985. 'Complementation' in Timothy Shopen (ed.), *Language Typology and Syntactic Description*, 3 vols. Cambridge University Press, vol. II, pp. 52–150.

Philippaki-Warburton, Irene and Catsimali, Georgia 1997. 'Control in Ancient Greek', in Gaberell Drachman, Angeliki Malikouti-Drachman, Ioannis Fykias and Celia Klidi (eds.), *Greek linguistics* 1995. *Proceedings of the 2nd international conference on Greek linguistics*, 2 vols. Graz: W. Neubauer Verlag GmbH., vol. II, 577–88.

Philippaki-Warburton, Irene and Catsimali, Georgia 1999. 'On Control in Greek', in Artemis Alexiadou, Geoffrey Horrocks and Melita Stavrou (eds.), *Studies in Greek Syntax.* Dordrecht: Kluwer Academic Publishers, pp. 153–68.

Porter, Stanley and O'Donnell, Matthew 2010. 'Building and Examining Linguistic Phenomena in a Corpus of Representative Papyri', in Trevor V. Evans and Dirk D. Obbink (eds.), pp. 287–311.

Rijksbaron, Albert 2006. *The Syntax and Semantics of the Verb in Classical Greek*, Third Edition. University of Chicago Press.

Robertson, Archibald T. 1919. *A grammar of the Greek New Testament in the light of historical research*, Third Edition. New York: George H. Doran.

Rowlandson, Jane 1998. *Women and society in Greek and Roman Egypt. A sourcebook.* Cambridge University Press.

Sevdali, Christina 2013. 'Ancient Greek infinitives and phases', *Syntax* 16.4: 324–61.

Spyropoulos, Vassilios 2005. 'The Syntax of Classical Greek Infinitive', in Katalin É. Kiss (ed.), *Universal Grammar in the Reconstruction of Ancient Languages*. Berlin: Mouton de Gruyter, pp. 295–337.

Spyropoulos, Vassilios 2007, 'Finiteness and Control in Greek', in William D. Davies and Stanley Dubinsky (eds.), *New Horizons in the Analysis of Control and Raising.* Dordrecht: Springer, pp. 159–83.

Stolk, Joanne and Nachtergaele, Delphine 2016. 'Dative for Accusative Case Interchange in Epistolary Formulas in Greek Papyrus Letters', *Symbolae Osloenses* 90.1: 1–42.

Thompson, Dorothy J. 1988. *Memphis under the Ptolemies*. Princeton University Press.

Vassiliou, Marina 2012. 'Measuring the Greek Infinitive Through Time', in Zoe Gavriilidou, Angeliki Efthymiou, Evangelia Thomadaki and Penelope Kambakis-Vougiouklis (eds.), *Selected papers of the 10th ICGL*, Komotini, Greece: Democritus University of Thrace, pp. 592–602.

Vierros, Marja 2021. 'Idiolect in focus: Two brothers in the Memphis Sarapeion (II BCE)', in Klaas Bentein and Marl Janse, *Varieties of Post-classical and Byzantine Greek*. Berlin: Mouton de Gruyter, pp. 39-74.

Wakker, Gerry C. 1988. 'Purpose Expressions in Homer', in Albert Rijksbaron, Henk A. Mulde and Gerry C. Wakker (eds.), *In the footsteps of Raphael Kühner.* Amsterdam: Gieben, pp. 327–44.

White, John L. 1986. *Light from ancient letters.* Philadelphia: Fortress Press.

Wilcken, Ulrich 1927. *Urkunden der Ptolemäerzeit* (Ältere Funde), 2 vols. Berlin – Leipzig: Walter de Gruyter, vol I: *Papyri aus Unterägypten.*

5. A Bilingual Scribe in Early Roman Tax Receipts from Elephantine

Ruth Duttenhöfer

This Chapter exploits a dossier of early Roman tax receipts which can be traced to a single scribe through his orthographic, linguistic, palaeographic and formulaic idiosyncrasies. Analysis of his scribal habits proves him to have been of Egyptian origin, writing both Demotic and Greek and probably speaking Greek as a second language.

In the tax collection system of Elephantine, situated at the southern border of Egypt, the transition from Ptolemaic to Roman rule is clearly marked by new tax titles, new types of receipts and new collectors.[1] In the third century BCE, Egyptian collectors writing Demotic and Greek collectors had worked alongside each other, sometimes independently, other times in collaboration.[2] In the second and first centuries BCE, however, a long breach in the Demotic documentation becomes apparent. For two centuries, tax collection in the island of Elephantine was exclusively handled by the Greek Bank at Syene and its bankers.[3]

Demotic tax receipts reappear from the time of Augustus after the Romans reorganised the Greek tax system in the region;[4] these are in fact the earliest Roman receipts from Elephantine so far, and they attest to two new types of taxes, *'pe* (lit. 'head'), the Demotic term for the poll tax, and *tny* ('tax'), possibly a tax on trades. By the end of Augustus' reign, Greek receipts appear employing

---

[1] An overview of Greek and Demotic tax receipts from Elephantine is found in Locher (1999: 297–317). A classification of the formulae of the receipts has already been made by Wilcken (1899: 58–129; esp. 118–27). Lists of collectors have been compiled by K. A. Worp in *O.Cairo GPW*, Appendice II, Esattori di tributi e loro assistenti a Syene/Elephantine, pp. 131–46.

[2] For the bilingual nature of Ptolemaic administration cf. Clarysse and Thompson (2006: 6–7).

[3] Hoffmann (2013); Duttenhöfer (2013).

[4] O.Mattha 122 (Aug. 11, guardian tax), O.Mattha 86 (Aug. 15, *tny pr-'3*) O.Mattha 171 (Aug. 18, *tny s n sm*); Brook.Cat. 68 (Aug. 30, *ḥḏ n 'pe.t*); O.Mattha 121 (Aug. 38, guardian tax); O.Mattha 194 (Aug. 39, *tny*). O.Eleph. inv. 1619 (Aug. 40, *'pe*). Hoffmann (2013: 96) put forward a possible explanation for the reappearance of Demotic script in tax receipts under Roman rule. Since Demotic is only found in the early phases of foreign rules, he suggests that the use of Demotic for matters of state administration reflects the hope of the new overlords that Egyptians would better cooperate if they were allowed to use their own script.

for the first time the Greek standard terms λαογραφία (poll tax) and χειρωνάξιον (trade tax); both taxes were the most important sources of money income for the state during the first century CE. From Tiberius onwards the major part of tax collection seems, again, to have been handled mostly in the Greek language; Demotic writing is represented by only a few scattered Demotic or bilingual texts. Greek becomes predominant, but this does not mean that Egyptians were not active in the tax collecting business. Observations on language and orthography will prove that at least one Egyptian scribe quickly adapted to writing Greek tax receipts.

Orthography in the highly formulaic texts[5] is, on the whole, consistent. There is hardly any room for variant spellings, except in the rendering of Egyptian names. Here one finds occasionally some variation, which can be attributed to the scribes struggling with Greek equivalents of Egyptian names.

It is curious, then, to find, in a relatively short period of time, several instances of irregular spellings for the ubiquitous term *laographia*, especially from Tiberius up to Nero. I have collected the following variations: λαγραφία, λαυγραφία, λαουγραφία, λευγραφία and λωγραφία.[6] Other examples of these variations can be found incidentally in texts from all over Egypt. Instances found via the Duke Databank of Documentary Papyri (accessible via papyri.info), are spread over the first and second centuries CE and beyond. Each isolated example might be a case of incidental misspelling, of omission or individual contraction, a misprint or a wrong reading. Alternatively, a variant may reflect the linguistic background of an individual scribe.[7]

With comparative material at hand, I took a closer look into the pheno-menon amongst the ostraca from Elephantine. From the published receipts of the early Roman Period, I traced four examples of the spelling λαγραφία to a single scribe, here called Scribe X,[8] whose name is not given in the receipts, but whose texts can easily be recognised by his idiosyncratic handwriting (cf. Figs. 1–7, below).

---

[5] διαγεγράφηκεν – name – ὑπὲρ – tax of year x – imperial titulature – amount [– day – subscript].

[6] λαγραφία: O.Leid. 178, O.Wilck. 6, O.Wilck. 12, CPR 10, 34, SB VI 9604, 13, SB 1, 1097, O.Berl. 23; λαυγραφία: SB 5, 7589; λαουγραφία: O.Berl. 24; λευγραφία: O.Wilck. 1239; λωγραφία: O.Wilck. 10. Cf. Gignac 1976: 301–2, and Gignac 1981: 31, who explains the variant forms as contraction.

[7] Similar examples are described in Dahlgren (2017: 71 and 156; 138–40 on language attitudes of Egyptian scribes).

[8] O.Leid. 178, O.Wilck. 6, SB 6, 9604,13 and O.Berl. 23.

Other unique spellings can be attributed to this scribe. Reexamination of the photographs of the ostraca reveals that the texts' previous editors sometimes corrected the original without noting the nonstandard spellings which are typical of this scribe. From the revised editions, found in the appendix at the end of this Chapter, his linguistic profile can be recognised by a series of spellings that occur in all his texts (Table 1).

Table 1. Spellings of Scribe X.

Scribe X	standard
δια<u>κ</u>εγράφηκε	διαγεγράφηκεν
λ<u>α</u>γραφίας	λαογραφίας
Τ<u>ει</u>βερίου	Τιβερίου
Γα<u>εί</u>ου	Γαίου
Σεβαστ<u>ός</u>, Σεβασ<u>δός</u>	Σεβαστοῦ
δεκαδ<u>ίου</u>	δεκαδύο
<u>δ</u>έσσαρες	τέσσαρες

These features strongly suggest that Scribe X is an Egyptian scribe writing Greek as a second language.[9] The evidence is as follows: (1) the typical confusion of voiced and voiceless stops (δια**κ**εγράφηκε, **δ**έσσαρες, Σεβασ**δ**ος), (2) itacistic writings (Τ**ει**βερίου / Γα**εί**ου, δεκαδ**ί**ου), (3) problems with the rendering of /o/ (ο or ου), (4) incorrect use of cases or declensions (Πελαιας for Πελαίου, cf. SB 6, 9604,13 in the following table, Σεβαστος for Σεβαστοῦ), and (5) the lack of Greek case endings of Egyptian proper names. Moreover, in writing Egyptian names the scribe displays occasionally genuine etymological knowledge of the underlying Egyptian form. For example, in O.Eleph. inv. 2175 (cf. Table 2), he writes Ουοννωφρι for *Wn-nfr* (Onnophris) and perhaps Πετουσιρις for *P3-dỉ-Wsỉr* (Petosiris), if this is not another instance of confusing ο and ου.[10]

I have found several more receipts in the same handwriting among the unpublished Elephantine ostraca, which all exhibit the above-mentioned spelling

---

[9] Cf. Gignac (1976: 46–8). Horrocks (2010: 112). Vierros 2012 for an analysis of an archive of bilingual notaries of Ptolemaic Egypt, pp. 139–173 for incorrect cases/declensions. Dahlgren (2016: 93–8) for a concise description of the typical features of Greek in Egypt with further literature and a history of research in this field on pp. 90–93.

[10] Other examples of etymologic spellings are e.g. P.Count. 35, 28: Οὐσόρσουβκις – *Wsỉr-Sbk*, standard Ὀσορσουχος; CEML 413 A2: Ουσωργηρις – *Wsỉr-wr*, standard Ὀσοροηρις; SB 18, 13202: Γορσενοῦφιος – Wrš-nfr, standard Ὀρσενουφις.

peculiarities. When one orders the texts chronologically, slight changes in the writing habits of Scribe X reveal a development in parts of the formula of his tax receipts (Table 2).

Table 2. Texts belonging to the hand of Scribe X.

text[11]	date/year	month	changing writings		names of taxpayers
O.Cairo GPW 63*	Tiberius 11	Hath. +Pham.	[λαγρα]φιας	Σεβαστος	-]βι Βιηγχι
O.Eleph. inv. 2175	Tiberius 14	-	λαγραφιας	Σεβαστος	Πετουσιρις Ουοννωφρι
OGL 1133	Tiberius 15	Paun.	-	-	-
O.Wilck. 6*	Tiberius 16	Pham.	λαγραφιας	Σεβαστος	Ζμηθις Παχνουβι
O.Leid. 178*	Tiberius ?	Pach.	λαγραφιας	Σεβαστος	Ζμηθι Παχνου[βι
OGL 1144	Tiberius ?	-	λαγραφιας	Σεβαστος	Πετ]εσουχος Πελαι()
O.Eleph. inv. 3680	Tiberius ?	-	λα[γραφιας]	[Σεβαστος]	Κτε[?
OGL 53	Tiberius ?	-	λαγρα|[φιας]	Σεβαστος	Πετορ[ζμηθι Πα]χνουβι
O.Berl. 23*	Tiberius 17	-	λαγρα()	Σεβαστος	Παχομπαζμηθι Ζμην
OGL 2081	Tiberius 17	-	λα()	Σεβαστος	Ψενπουηρι
OGL 202	Tiberius 17	-	λαγρα()	Σεβαστος	Πετεχνουβι Βιηνχ
O.Eleph. inv. 4317*	Tiberius 19	-	λαγρα()	Σεβασδος	Πετορζμηθι Παων`νω?
OGL 838*	Tiberius 19	-	χειρωναξιου	Σεβασδος	Πετορζ]μηθι
SB 6, 9604,13*	Gaius 2	-	λαγρα()	Σεβασδος	Παχνουβι Πελαιας
OGL 975	Gaius 2	-	λαγρα()	Σεβασδος	Πελαιας Πετεσουχο`ς´
OGL 791	Gaius ?	-	λαγρ()	Σεβασδος	-

First, the indication of month and day at the end of the receipt occurs only during the first period from Tiberius' years 11–16; later, it is abandoned. Second, during the same period, the term λαγραφία is written out in full, but from year 17 onward, the scribe abbreviates the term. On this basis, I have inserted the texts whose exact dates are lost after year 16 of Tiberius: O.Leid. 178 was dated to the month Pachon; OGL 1144 preserves the full tax title, and in O.Eleph. inv. 3680 and OGL 53, the full tax title is expected to be written out in the *lacuna* because of the space available. Finally, a change occurs in the imperial titulature of the

---

[11] Texts marked with * have been revised or edited in the appendix at the end of this Chapter. Unpublished texts of the Louvre (OGL) and of the joined excavations at Elephantine of the German Archaeological Institute Kairo and the Swiss Institute for Architectural and Archaeological Research on Ancient Egypt (O.Eleph. inv.) will be published by me. For abbreviations of editions of papyri and ostraca, see the *Checklist of Greek, Latin, Demotic and Coptic Papyri, Ostraca and Tablets* (available online at papyri.info/docs/checklist).

dating formula whereby the title Σεβαστος changes to Σεβασδος (always in the wrong case, *l.* Σεβαστοῦ).[12]

Usually, the Egyptian proper names of the taxpayers do not take a Greek ending and are not abbreviated. Yet, the nominative of a few names has occasionally been written out correctly, as in Πετουσιρις, Πετεσουχος, Πελαιας and once in Ζμηθις. The genitive of the patronymic is never realised. Scribe X uses basically only one undeclined form for each name.[13]

To the linguistic profile emerging in the orthography, one can add the assessment of the scribe's handwriting. The hand is seemingly clumsy but betrays that this scribe had actually had ample practice. His handwriting does not reflect a scribe's training, which is always connected with the reproduction of a certain style. Scribe X knew the form of each single Greek letter but never learned to produce cursive writing even though, with time, his hand became noticeably more fluent. One can speculate on the model he had in mind for writing the Greek letters. In most cases, it seems to have been epigraphical forms, certainly for *alpha* and *epsilon*. Most letters are written separately, each formed by one, two or three strokes. In only two cases, letters are almost naturally connected by way of a ligature. In the combinations -αρ- and -θι-, the middle bar of the preceding letter leads into the vertical stroke of an adjacent *rho* or *iota*. Intriguing is the form of *kappa*. It regularly has a horizontal connecting element at its upper right end ( ), which is not part of the archetype Κ. Scribe X may have adopted the form from late Ptolemaic documents written in a cursive script, which have similar connecting strokes.

In conclusion, Scribe X was *not* well trained as a professional Greek scribe, but he could read Greek and reproduced the letters on a privately formed basic knowledge of the Greek alphabet; he even corrected his own writing by adding a forgotten syllable or letter above the line, cf. O.Leid. 178, 1 (see (3) below); O.Eleph. inv. 4317, 2 (see (5) below); OGL 975 (see Table 2). The evidence suggests that he understood and probably spoke Greek. He reproduced the correct wording of a Greek receipt, but his spellings and orthography are not standard. His bilingualism is most evident by the fact that he sometimes spells Egyptian names phonetically.

Further clues to the identity of this scribe can be extracted from the subscripts to the receipts which are found in seven out of sixteen of Scribe X's texts, known to me thus far (cf. Table 2).

---

[12] See Dahlgren (2017: 67, 83–4).

[13] Cf. Pestman (1993: 485–96), for Greek endings and declensions of Egyptian personal names.

Even though most tax receipts of the early first century CE do not regularly reveal information about the scribes or the tax collectors, a considerable amount of documents show signs of official handling. There are two types of Greek subscripts, depending on the number of officials involved in the collection process.[14]

Type I subscripts are defined by the fact that the subscript and the body of the receipt are written in the *same hand*. The subscript always states the name of the writer, mostly in the form *N.N.* ἔγραψα / ἔγραψεν, 'I, *N.N.*, have written' or '*N.N.* has written.'[15]

This form of subscript is actually a signature of the collector that verifies/ corroborates the content of the receipt. His signature was not obligatory nor was it necessary, since the collector himself was writing and handing out the receipt in person. Most *praktores* were acting alone, but we know of two or more *praktores* working at the same period in the same district.[16]

Type II subscripts are written by a *different hand* than the receipt. They always take the form: *N.N.* ἐπηκολούθηκα, '*N.N.* has followed.'[17]

A subscript with the verb ἐπακολουθέω acknowledges that a second official was present and presumably checked the payment. The 'follower', in my view, is a controller on the same level as the *praktor* who wrote the receipt, but with a different function.[18] To construct any hierarchy between the two persons involved in the collection would be to adopt a later second-century practice,[19] clearly different from the situation in the first century.

Three receipts of Scribe X have been subscribed in the manner of type II, which indicates that occasionally a controlling instance was involved in the collecting process.

---

[14] Cf. Duttenhöfer (2022: 364).

[15] *N.N.* ἔγραψα: O.Wilck. 12; O.Wilck. 18; SB 6, 9545, 6. This type is the standard form for signatures in the first century from the time of Vespasian onwards. *N.N.* ἔγραψε: O.Cairo GPW 62, SB 1, 1097 (with BL 8, 305), CPR 10, 34. (The variation between ἔγραψα and ἔγραψε could be phonetic in nature, cf. Leiwo (2010: 114–18). In O.Wilck. 20, the *praktor* signs with διὰ Ἑρμωνᾶτος; in SB 6, 9545, 1, two *praktores* just give their names and the title.

[16] Cf. Duttenhöfer (2022: 364-367).

[17] Cf. the examples SB 20, 15044, O.Leid. 177, O.Wilck. 7; cf. O.Wilcken I, pp. 76–7.

[18] Cf. Duttenhöfer (2022: 367-370).

[19] In the second century collectors (*misthotai*, *epiteretai* and even *praktores*) use clerks or assistants (with the titles γραμματεύς or βοηθός) to write the receipts and probably handle the collection process, cf. Reiter (2004: 123). The subscript of the collector appears, if at all, at the end, often in the form *N.N.* σεσημείωμαι.

text	year	subscripts in second hand
O.Eleph. inv. 3680	Tiberius ?	[Πελαί]ας Ἀρίστω(νος) [ἐπηκολ]ού(θηκα)
SB 6, 9604, 13	Gaius 2	Ἀρίστων Πελαίου ἐπεκλούθ(ηκα) (*l.* ἐπηκολούθηκα)[20]
OGL 791	Gaius ?	Ἀρίστων [Πελαίου ἐπεκλούθ(ηκα)] (*l.* ἐπηκολούθηκα)

However, four texts of Scribe X contain a Demotic subscript which was written in all instances by the same Demotic hand.

text	year	subscripts in Demotic
O.Cairo GPW 63	Tiberius 11	*ḫ P3-dĭ-Wsĭr-ns-mtr ...* 'Petosorsmethis has written ...'
O.Leid. 178	Tiberius ?	*sḫ P3-dĭ-Wsĭr-ns-mtr ...* 'Petosorsmethis has written ...'
O.Eleph. inv. 4317	Tiberius 19	*ḫ3t-sp 19 P3-dĭ-Wsĭr-ns-mtr* Year 19, Petosorsmethis'
OGL 838	Tiberius 19	*ḫ3t-sp 19 P3*[*-dĭ-Wsĭr-ns-mtr*]'Year 19, Petosorsmethis'

The Demotic subscripts attested in O.Cairo GPW 63 and O.Leid. 178 can be compared with the official signatures of scribes/collectors in purely Demotic receipts, where both receipt and signature, originate from the same hand. The Demotic scribes identify themselves using the formula *sḫ N.N.*, '*N.N.* has written'.[21] Thus, the Demotic signatures have the same function as type I subscripts in the Greek receipts, acknowledging the content of the receipt with the signature of the scribe. For O.Cairo GPW 63 and O.Leid. 178, therefore, it is inevitable to conclude that the Demotic scribe Petosorsmethis has produced the receipt in Greek and signed it in Demotic. For the two later examples O.Eleph. inv. 4317 and OGL 838, this interpretation may not be equally compelling, since the formula is slightly different as the element *sḫ* 'has written' is missing.[22] Yet, seen in context with the former subscripts, they ought to be understood in the same sense as O.Cairo GPW 63 and O.Leid. 178.

It is almost impossible to prove that the same hand wrote Demotic and Greek in the same text, when no indication in the context to that effect has been made. Editors generally take a Demotic subscript to a Greek receipt as deriving from a second person, and there are only few examples of a scribe who did write in both languages in the same document.[23]

---

[20] The names Pelaias (Eg.) and Ariston (Gr.) indicate the Graeco-Egyptian background of the collectors and suggest that Pelaias, son of Ariston, and Ariston, son of Pelaias, were father and son.

[21] Cf. the examples of Demotic Ostraca O.Louvre 243, 584, 704.

[22] Cf. Appendix (5) O.Eleph. inv. 4317, note on l. 7.

[23] Only Petepiphis, a tax collector from the third century BCE in Elephantine, writes both Greek

In our case, I suppose – judging from the linguistic profile – that the writer of the Greek text probably has an Egyptian background. From the fact that four texts display type I subscripts in Demotic, I conclude that Petosorsmethis was, himself, the *praktor* who wrote the receipt in Greek and signed it with his name in Demotic.

To strengthen this hypothesis, I observe that both parts (Greek and Demotic) of the four receipts are written with a *kalamos*[24] and probably in each case with the same *kalamos*, since the thickness of the strokes is equal. The ink flow and the colour of the ink look the same throughout for each receipt; in addition, the line spacing between the Greek and the Demotic text parts seems regular.

There is little doubt that the Demotic scribe Petosorsmethis is the same as Scribe X who wrote the Greek parts of the receipts. He was not only bilingual, but also a *digraphos*. As a professional Demotic scribe, he was fast assimilating to the task of writing Greek tax receipts. He is evidence of the inclusion of indigenous scribes in the Roman tax collecting process.

---

and Demotic, but very rarely in the same document, cf. S. P. Vleeming, O.Varia 14, p. 41 note (ff) and p. 47 note (xx). O.Varia 14 seems to be the only example where he wrote both the receipt in Demotic and the subscript in Greek (compare the subscript to O.Varia 17). Cf. Muhs (1998: 74).

[24] The *kalamos* is the writing pen used in third-century BCE tax receipts by Greek scribes, whereas Demotic scribes used a reed brush. From the second century BCE onwards, Egyptians used a Greek *kalamos* for writing Greek, cf. Clarysse (1993).

Appendix

Five published receipts written by the Scribe Petosorsmethis have been revised and corrected: (1) - (4) and (7). In addition, two new bilingual receipts from his hand are published in (5) and (6).

(1)	O.Cairo GPW 63	24–25 CE	Fig. 1
(2)	O.Wilck. 6	30 CE	Fig. 2
(3)	O.Leid. 178	25–30 CE	Fig. 3
(4)	O.Berl. 23	30–31 CE	Fig. 4
(5)	O.Eleph. inv. 4317	32–33 CE	Fig. 5
(6)	OGL 838	32–33 CE	Fig. 6
(7)	SB 6, 9604, 13	37–38 CE	Fig. 7
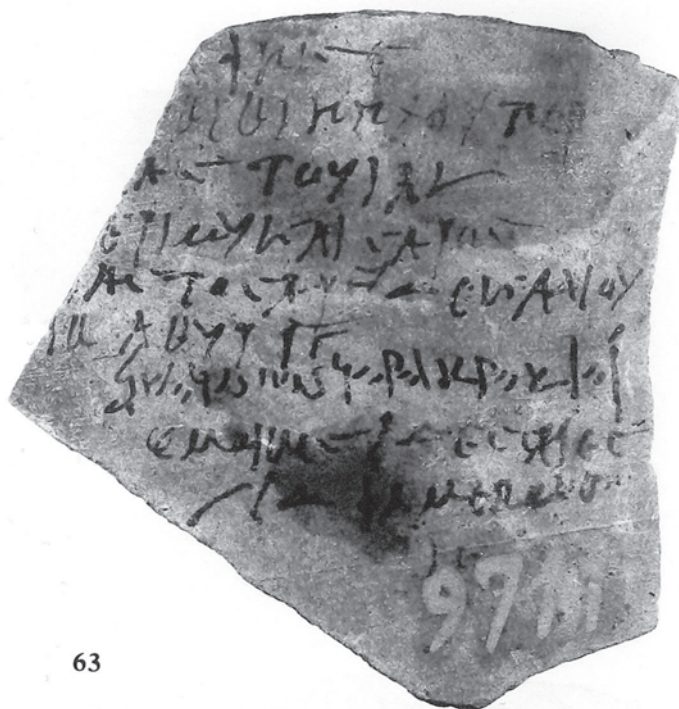
Figure 1. O.Cairo GPW 63 (Cairo Museum CG 9711), Scan of Tav. XV.

(1) O.Cairo GPW 63 (9 November 24 CE and 11 March 25 CE)

[διακεγρ]άφηκε
[*ca.* 6   ]︲βι Βιηνχι ὑπὲρ
[λαγραφ]ίας τοῦ ια (ἔτους)
[Τειβ]ερίου Καίσαρος
[Σε]βαστος ἀργυ(ρίου) (δραχμὰς) δέκα δίου,   5
[(γίνονται) (δραχμαὶ)] ιβ. Ἀθὺρ ι̅γ̅.
Dem.   *sẖ P3-dỉ-Wsỉr-ns-mty s3 P3-*︲︲︲ *(?) ẖr sttr 3.t*
ὁμοίως (δραχμὰς) δέσσαρες,
(γίνονται) (δραχμαὶ) δ. Φαμενὼθ
     ι̅ε̅.   10

───────────

1 *l.* διαγεγράφηκε | 2 *l.*]︲βις Βιήγχιος | 3 *l.* λαογραφίας, Ⳑ | 5 *l.* Σεβαστοῦ,
αργυ^υ ϛ, *l.* δύο | 8 ϛ, *l.* τέσσαρας | 9 / ϛ

Has paid ---bis, son of Bienchis, for *laographia* of the 11th year of Tiberius Caesar Augustus, twelve silver dr., = 12 dr., Hathyr 13.

(*Demotic*) Has written Petosorsmetis, son of P---, concerning 3 staters.

Ditto four dr., = 4 dr. Phamenoth 15.

1 [διακεγρ]άφηκε: *ed. pr.* [διαγεγρ]άφηκε. The introducing verb in this scribe's hand is always spelled with *kappa,* confusing voiced *gamma* with voiceless *kappa*, and always without *nu* at the end. I restore in the *lacuna* the form expected from this scribe. The form διαγεγράφηκεν is exclusively used as the opening formula in Elephantine tax receipts from 16–67 CE, cf. O.Leid. 178, note on l. 1.

2 [7] ̣βι Βιήνχι: *ed. pr.* [7] ̣βι() Βιήνχι(ος); no sign of abbreviation can be seen.

3 [λαγραφ]ίας: *ed. pr.* [λαογραφ]ίας; I restore the scribe's spelling.

4 [Τειβ]ερίου: *ed. pr.* [Τιβ]ερίου. Τειβερίου is always written with itacistic ει for ι in this scribe's hand.

7 *P3-dỉ-Wsỉr-ns-mty s3 P3-* ̣ ̣ ̣ *(?).* The editors of O.Cairo GPW 63 have recognised, that the Demotic subscript was written by the same hand as the one in O.Leid. 178. What followes the name of the scribe Petosorsmethis, should be the remains of his father's name. But the reading proposed in O.Leid. 178, 6: *Pa-Wn-nfr* could not be verified here.

8 δέσσαρες: Confusion between voiced and voiceless stops: δ/τ.

Figure 2. O.Wilck. 6 (OGL 1554) © Scan: Musée du Louvre.

(2) O.Wilck. 6 (15 March 30 CE)

διακεγράφηκε Ζμῆθις
Παχνουβι ὑπὲρ λαγραφίας
τοῦ ις (ἔτους) Τειβερίου
Καίσαρος Σεβαστος
ἀργυ(ρίου) (δραχμὰς) ὀκτώ, (γίνονται) (δραχμαὶ) η.   5
   Φαμεν(ὼθ)   (vac.) ιθ.

---

1 *l.* διαγεγράφηκε | 2 *l.* Παχνούβιος, λαογραφίας | 3 L, *l.* Τιβερίου | 4 *l.* Σεβαστοῦ | 5 αργ^υ ϛ, /ϛ

Has paid Zmethis, son of Pachnubis, for *laographia* of the 16th year of Tiberius Caesar Augustus, eight silver dr., = 8 dr., Phamenoth 19.

1 διακεγράφηκε: *ed. pr.* διακεγράφηκεν; cf. (1) O.Cairo GPW 63, note on l. 1

2 Παχνουβι: For the father's name, the *apparatus* of the *editio princeps* suggests Παχνούβι<ος> with pointed brackets, as if the ending was inadvertently left out by the scribe. But comparison with the other examples shows that the scribe only occasionally writes nominative endings for Egyptian names; in most cases, he leaves names without endings.

3 τοῦ ις (ἔτους): *ed. pr.* τοῦ ι<γ> L. While the reading ις L was originally found in 'Fröhners Papieren', from where U. Wilcken took over the transcript without having seen the ostracon, the 13th year was conjectured by him (cf. O.Wilck. 6, note on l. 3) because the same taxpayer had paid the full amount for *laographia* of the 16th year in O.Wilck. 8. But there, in ll. 3–4, not the 16th but the 10th year is written: instead of ἐκ|(και)δεκάτου (ἔτους), one has to read τοῦ | δεκάτου (ἔτους), cf. Taf. I. a. in Wilcken 1899.

4 Σεβαστός: *ed. pr.* Σεβαστοῦ; the scribe is not aware of the right declension type for Σεβαστός.

5 ἀργυ(ρίου) ϛ: *ed. pr.* δραχ(μάς).

6 Φαμεν(ὼθ) ιθ: *ed. pr.* Φαμενώθ. There is a gap between Φαμεν and ιθ; the numeral at the end is clearly marked with a horizontal stroke above. The receipt dates from 15th of March 30 CE.
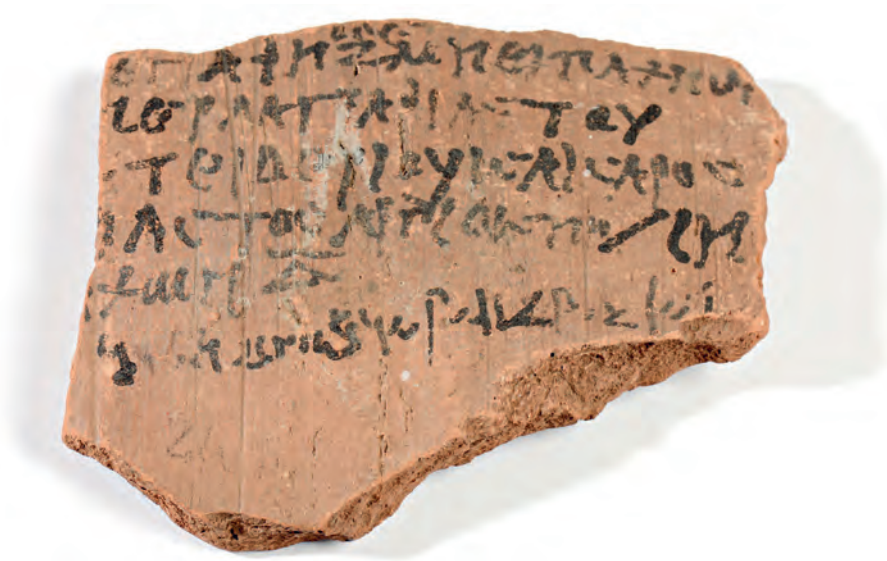
Figure.3. O.Leid. 178 (Leiden Museum AT 85d) © National Museum of Antiquities, Leiden, The Netherlands.

(3) O.Leid 178 (= O.Wilck. 1322) (29 April 25–30 CE)

[διακ]εγράφη‵κε′ Ζμηθι Παχνου-
[βι ὑ]πὲρ λαγραφίας τοῦ
[. .] (ἔτους) Τειβερίου Καίσαρος
[Σε]βαστος ἀργυ(ρίου) (δραχμὰς) ὀκτώ, (γίνονται) (δραχμαὶ) η.
[Π]αχὼν δ.   5
Dem.   *sḫ P3-di-Wsir-ns-mty s3 P3- . . . (?) ḫr sttr 2.t*
_____

1 *l.* διαγεγράφηκε Ζμῆθις Παχνού|βιος | 2 *l.* λαογραφίας | 3]Ḷ, *l.* Τιβερίου |
4 *l.* Σεβαστοῦ, αργ^υ ϛ, /ϛ
Has paid Zmethis, son of Pachnubis, for *laographia* of year ? of Tiberius Caesar
Augustus eight silver dr., = 8 dr., Pachon 4.
(Demotic) Has written Petosorsmetis, son of P---, concerning 2 staters.

1 [διακ]εγράφη῾κε′: *ed. pr.* [διαγ]εγράφη῾κε′.

   Ζμηθι: *ed. pr.* Ζμῆθι(ς). The name Zmethis is written without ending; there is no sign of abbreviation.

1–2 Παχνου|[βι ὑ]πέρ: *ed. pr.* Παχνού|[βιο(ς) ὑ]πέρ. Probably the name was written without a Greek ending. The *lacuna* at the left side is gradually getting smaller, allowing according to the formula in l. 1 four letters, in l. 3 and 4 two letters, in l. 5 one letter.

3 [ˌ ˌ] (ἔτους): *ed. pr.* [ˌ(ˌ) ἔτου]ς; the word ἔτους cannot be written out in full in the *lacuna*. I take the traces as belonging to the expected ἔτος sign (L) protruding its horizontal out of the *lacuna*.

5 [Π]αχὼν δ: *ed. pr.* [(ἔτους) ˌ ˌ Π]αχὼν δ. There is only π and part of α missing at the beginning of the line, the scribe never repeats the year at the end of the receipt. The expected line above δ marking the numeral is missing here.

   I date the receipt because of the similarities to O.Cairo GPW 63 and the formula employed from year 11–16 of Tiberius, cf. the 'Table 2: Texts belonging to the hand of Scribe X'.

6 *sḫ P3-dỉ-Wsỉr-ns-mty s3 P3-ˌ ˌ ˌ (?)*: *ed. pr. P3-dỉ-Wsỉr-ns-mty s3 Pa-Wn-nfr*. The editors of O.Cairo GPW 63 have recognised, that the Demotic subscript was written by the same hand as the one in O.Leid. 178. But the reading of the father's name *Pa-Wn-nfr*, 'son of Paonnophris', originally proposed in O.Leid. 178, 6, could not be verified in O.Cairo GPW 63, 7: *P3-ˌ ˌ ˌ (?)*. The father's name remains incertain.
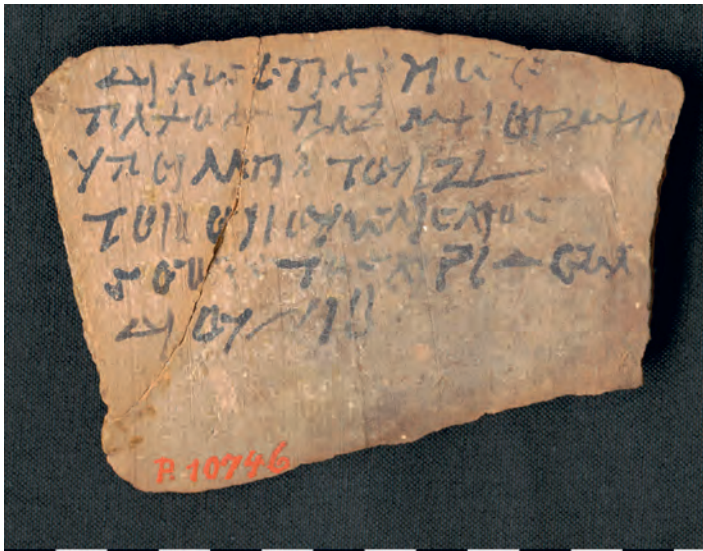
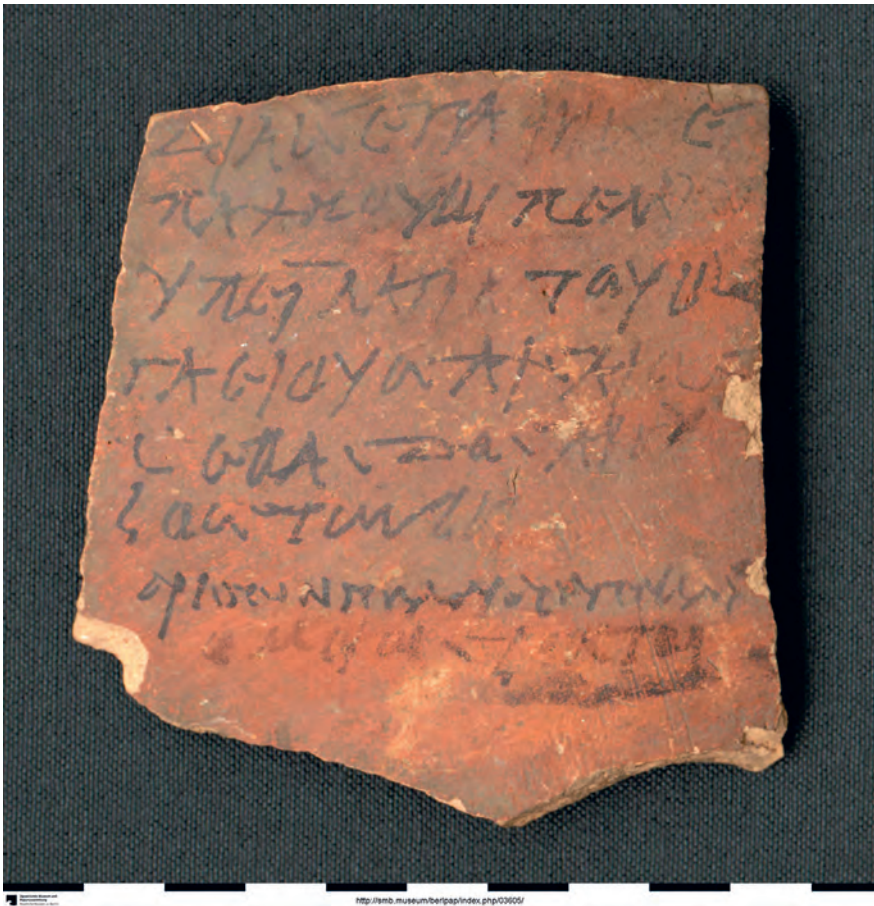

Figure 4. O.Berl. 23 © Staatliche Museen zu Berlin - Ägyptisches Museum und Papyrussammlung, Scan: Berliner Papyrusdatenbank P 10746.

(4) O.Berl. 23* (30–31 CE)

διακεγράφηκε
Παχομπαζμηθι Ζμην
ὑπὲρ λαγρα(φίας) τοῦ ιζ (ἔτους)
Τειβερίου Καίσαρος
Σεβαστος ἀργυ(ρίου) (δραχμὰς) δέκα   5
δίου, (γίνονται) (δραχμαὶ) ιβ.

_____

1 *l.* διαγεγράφηκε | 2 *l.* Παχομπάζμηθις Ζμήνιος(?) | 3 λαγρ͞α, *l.* λαογρα(φίας);
L | 4 *l.* Τιβερίου | 5 *l.* Σεβαστοῦ, αρ͞γ ϛ | 6 *l.* δύο, /ϛ

Has paid Pachompazmethis, son of Zmenis(?), for *laographia* of year 17 of
Tiberius Caesar Augustus, twelve silver dr., = 12 dr.

* A reedition of this text which already included some of the following corrections was
published by Reiter (2017).
2 Παχομπαζμηθι Ζμην: *ed. pr.* Παχωμπαζ(μῆθις?) μη(τρὸς) Θιζμήθι(ος). Both names
are written without Greek ending. The name Pachompazmethis is typical of Elephantine,
cf. Trismegistos *s.v.* 'Pasmet'; see now Jennes (2014: 149–51). Ζμην might be a variant
writing of the Egyptian name *Ns-Mn*, Greek Ἐσμῖνις, Ζμῖνις, Ζμῖν and Ζμῆνις, once
attested also for Elephantine in P.Äg. Handschriften 272 descr.; Reiter (2017: 295),
interprets Ζμην as a variant writing for Ζμενπῶς/Ζμεντπῶς, which is connected to
Ζμῆθις, *Ns-p3-mtr*, cf. Jennes (2014: 149).
3 λαγρα(φίας): *ed. pr.* λαογρα(φίας).
4 Τειβερίου: *ed. pr.* Τβιβερίου. Cf. Reiter (2017: 295).
5 ἀργυ(ρίου): *ed. pr.* Χ(οιὰκ) ͞γ. There is no indication of month and day here; the receipt
dates from year 17 of Tiberius, 30–31 CE.
6 δίου: *l.* δύο. υ was clearly felt as /i/, whereas /o/ was spelled as ου, cf. Dahlgren (2016:
91 and 93).
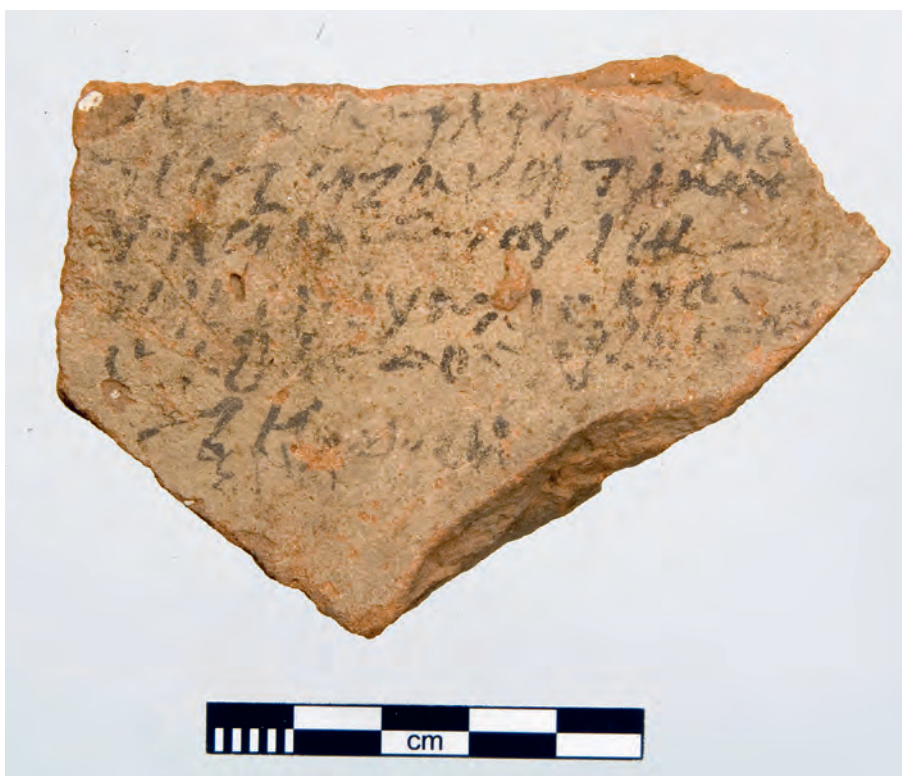
Figure 5. O.Eleph. inv. 4317. Excavations on Elephantine conducted by the German Archaeological Institute Kairo and the Swiss Institute for Architectural and Archaeological Research on Ancient Egypt © Deutsches Archäologisches Institut Kairo.

(5) O.Eleph. inv. 4317 (32–33 CE)

Surface partly abraded and damaged. Demotic read by S. P. Vleeming.

διακεγράφηκε
Πετορζμηθι Παων᾽νω´
ὑπὲρ λαγρα(φίας) τοῦ ιθ (ἔτους)
Τειβερίου Καίσαρος
Σεβασδος ἀργυ(ρίου) (δραχμὰς) ὀκτώ,   5
(γίνονται) (δραχμαὶ) η.
Dem.   ḥ3t-sp 19 P3-dĭ-Wsỉr-ns-mtr (sttr) 2.t

----

1 *l.* διαγεγράφηκε | 2 *l.* Πετόρζμηθις Παοννώ(φριος)? | 3 υπερ̄ λαγρᾱ, *l.* λαογραφίας L | 4 *l.* Τιβερίου | 5 *l.* Σεβαστοῦ, αργ⁰ ϛ | 6 / ϛ

Has paid Petorzmethis, son of Paonnophris, for *laographia* of year 19 of Tiberius Caesar Augustus, eight silver dr., = 8 dr.
(Demotic) Year 19, Petosorsmethis, 2 (staters).

7 The Demotic line starts on the same level as the Greek beginning of l. 6, but descends below that level, when the scribe realises that there is not enough space to complete his subscript. At the end, there is clearly written *2.t*, but *sttr,* 'staters', was left out and has to be understood.

Since the tax payer here and perhaps in (6) OGL 838 is called Petorzmethis, the Greek equivalent of Demotic Petosormethis, one might wonder if the Demotic subscript refers to the tax payer rather than to the tax collector who has accidentally the same name, which is the most common name in Elephantine, see Jennes 2014. I know indeed of one rare example of a Demotic subscript to a Greek receipt (Berliner Papyrussammlung P.10807, unpublished), which repeats the taxpayer's name in Demotic. In favour of this interpretation, one might even argue that (5) O.Eleph. inv. 4317 and (6) OGL 838 employ another formula than (1) O.Cairo GPW 63 and (3) O.Leid. 178, which therefore should have another function. However, even if one reads the subscripts of (5) O.Eleph. inv. 4317 and (6) OGL 838 in this sense, one cannot deny the fact that in all receipts the same Greek hand and the same Demotic hand are at work, which suggests in itself, that they might have been produced by the same scribe.
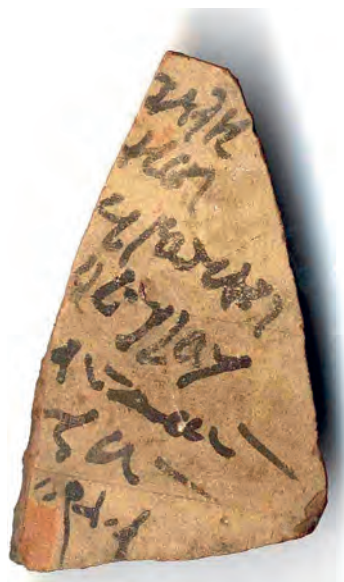
Figure 6. OGL 838 © Scan: Musée du Louvre.

(6) OGL 838 (32–33 CE)

Broken at left. Demotic read by S. P. Vleeming.

[διακε]γράφη-
[κε]μηθι
[ὑπὲρ χ]ειροναξί-
[ου τοῦ ιθ (ἔτους) Τει]βερίου
[Καίσαρος Σεβ]ασδος   5
[ἀργυ(ρίου) (δραχμὰς) εἴκοσι,] (γίνονται) (δραχμαὶ) κ.
Dem.   ḥ3t-sp 19 P3[-dỉ-Wsỉr-ns-mtr]

_____

1–2 *l.* διαγεγράφη|κε | 3–4 *l.* χειρωναξίου | 4 *l.* Τιβερίου | 5 *l.* Σεβαστοῦ | 6 / ϛ

Has paid ---methis, son of ---, for *cheironaxion* of year 19 of Tiberius Caesar Augustus, twenty silver dr., = 20 dr.
(Demotic) Year 19, Petosorsmethis.

2 The taxpayer's name might be Petorzmethis, which is the most common name in Elephantine, see Jennes (2014); the father's name is lost. Cf. (5) O.Eleph. inv. 4317, note on l. 7.

Figure 7. SB 6, 9604, 13 © Staatliche Museen zu Berlin – Ägyptisches Museum und
Papyrussammlung, Scan: Berliner Papyrusdatenbank P 12705.

(7) SB 6, 9604, 13 (37–38 CE)

διακεγράφηκε
Παχνουβι Πελαιας
ὑπὲρ λαγρα(φίας) τοῦ β (ἔτους)
Γαείου Καίσαρος
Σεβασδος ἀργυ(ρίου)   5
(δραχμὰς) ὀκτώ, (γίνονται) (δραχμαὶ) η.
m. 2   Ἀρίστων Πελαίου ἐπεκλούθ(ηκα).
m. 1   Ὁμοίως (δραχμὰς) ὀκτώ.

2 *l.* Πάχνουβις Πελαίου | 3 υπερ̄ λαγρā, *l.* λαογραφίας, L | 4 *l.* Γαίου | 5 *l.* Σεβαστοῦ | 6 𐅹, / 𐅹 | 7 επεκλου^θ, *l.* ἐπηκολούθ(ηκα)

Has paid Pachnubis, son of Pelaias, for *laographia* of year 2 of Gaius Caesar Augustus, eight silver dr., = 8 dr.
m. 2: Ariston, son of Pelaias, I have followed.
m. 1: Ditto 8 dr.

2 Παχνουβι Πελαιας: *ed. pr.* Παχνοῦβι(ς) Πελαίου. There is no sign of abbreviation at the end of Παχνουβι. The father's name has the nominative ending, not the genitive.
3 λαγρα(φίας): *ed. pr.* λα(ο)γρα(φίας). There is a superfluous horizontal mark above the end of υπερ̄, and again, barely visible, above λαγρā; but there is no reason for expanding to λα(ο)-, as we know from the parallels that this scribe always writes the term λαγρā.
4 Γαείου is written with itacistic ει for ι as in Τειβερίου.
5 Σεβασδος, *l.* Σεβαστοῦ: Confusion of voiced and voiceless stops, δ/τ.
7 ἐπεκλούθ(ηκα): *l.* ἐπηκολούθ(ηκα). For vowel loss between consonants (*syncope*), esp. in connection with liquids, see Gignac (1976: 306–10); on augmentation, esp. in compound verbs, see Gignac (1981: 233 and 248).

References

For the editions of papyri and ostraca, see the Checklist of Greek, Latin, Demotic and Coptic Papyri, Ostraca and Tablets (available online at papyri.info/docs/checklist).

Capasso, Mario, Davoli, Paola, and Pellé, Nastascia (eds.) 2022, Proceedings of the 29th International Congress of Papyrology. Lecce, 28 July-2 August 2019, 2 vols., University of Salento [http://siba-ese.unisalento.it/index.php/29th_ICP/issue/view/1881].

Clarysse, Willy 1993. 'Egyptian Scribes Writing Greek', *CE* 68: 186–201.

Clarysse, Willy and Thompson, Dorothy J. 2006. *Counting the People in Hellenistic Egypt*, 2 vols., Cambridge University Press.

Dahlgren, Sonja 2016. 'Towards a Definition of an Egyptian Greek variety', *Papers in Historical Phonology* (*PiHPh*, journals.ed.ac.uk/pihph) 1: 90–108.

Dahlgren, Sonja 2017. *Outcome of long-term language contact: Transfer of Egyptian phonological features on Greek in Graeco-Roman Egypt*, Helsinki, doctoral dissertation.

Duttenhöfer, Ruth 2013. 'Greek Ostraca from Elephantine', in Raue, Seidlmayer and Speiser (eds.), pp. 53–8.

Duttenhöfer, Ruth 2022. 'Quittungsformular, Schreibergepflogenheiten und Personalstruktur im Steuereinzug von Elephantine in frührömischer Zeit', in Capasso, Davoli and Pellé (eds.), vol. I, pp. 361–371.

Gignac, Francis T. 1976. *A Grammar of the Greek Papyri of the Roman and Byzantine Periods. I. Phonology.* Milan: Cisalpino - Goliardica.

Gignac, Francis T. 1981. *A Grammar of the Greek Papyri of the Roman and Byzantine Periods. II. Morphology*. Milan: Cisalpino - Goliardica.

Hoffmann, Friedhelm 2013. 'Demotic Ostraca from Elephantine', in Raue, Seidlmayer and Speiser (eds.), pp. 91–6.

Horrocks, Geoffrey C. 2010 [1997]. *Greek: A history of the language and its speakers*. Oxford: Wiley-Blackwell.

Jennes, Gwen 2014. 'Osiris-Zmithis in Egyptian Anthroponyms', *JJP* 44, pp. 137–53.

Leiwo, Martti 2010. 'Imperatives and Other Directives in the Greek Letters from the Mons Claudianus', in Trevor V. Evans and Dirk D. Obbink (eds.), *The Language of the Papyri*, Oxford University Press, pp. 97–119.

Locher, Josef 1999. *Topographie und Geschichte der Region am ersten Nilkatarakt in griechisch-römischer Zeit*, Stuttgart – Leipzig: Teubner.

Muhs, Brian 1998. 'The Chronology of the Reign of Ptolemy II Reconsidered',
     in A. Verhoogt and S. P. Vleeming (eds.), *The Two Faces of Graeco-Roman
     Egypt. Greek and Demotic and Greek-Demotic Texts and Studies Presented to
     P.W. Pestman*, Leiden: Brill, pp. 71–85.

Pestman, Pieter W. 1993. *The Archive of the Theban Choachytes (Second Century
     B.C.). A Survey of the Demotic and Greek Papyri Contained in the Archive*,
     Louvain, Peeters.

Raue, Dietrich, Seidlmayer, Stephan J. and Speiser, Philipp (eds.) 2013. *The First
     Cataract of the Nile. One Region – Diverse Perspectives*, Berlin – Boston: De
     Gruyter.

Reiter, Fabian 2004. *Die Nomarchen des Arsinoites*, Paderborn – München – Wien
     – Zürich: Schöningh.

Reiter, Fabian 2017. 'Lesevorschläge zu griechischen Steuerquittungen römischer
     Zeit aus Elephantine', *APF* 63, pp. 294–6.

Vierros, Marja 2012. *Bilingual notaries in Hellenistic Egypt. A study of Greek as a
     second language,* Brussels.

Wilcken, Ulrich 1899. *Griechische Ostraka aus Ägypten und Nubien. Ein Beitrag
     zur antiken Wirtschaftsgeschichte*, vol. I, München: Giesecke & Devriendt;
     2nd ed. Amsterdam: Hakkert.

6. Documentary Papyri as 'Multimodal' Texts. Aspects of Variation in the Nepheros Archive (IV CE)

Klaas Bentein

1 Introduction

Following Labov's seminal findings in the 1960s,[1] studies have highlighted and analyzed the importance of linguistic variation for the expression of social meaning, both synchronically and diachronically. From a synchronic point of view, linguistic variation enables social visibility, as Spitzmüller (2013: 1) writes:

> 'Weil wir sprachliche Variation haben, weil es verschiedene Möglichkeiten gibt, etwas (wiederum nur im Sinne einer semiotischen *Referenz*) zu *zeigen*, können wir durch Sprache zugleich auch *uns selber* zeigen. Kommunikative Varianten oder "Stile" sind daher ... wesentlich "ein Mittel zur Steigerung sozialer Sichtbarkeit" und mehr als das: Sie *ermöglichen* überhaupt erst "soziale sichtbarkeit".'

Studying the social import of speakers' and writers' choices helps us to better understand the message conveyed by ancient texts, or even to reconstruct their social context. With regard to documentary texts, for example, Krüse (2002: 879–894) has noted that the way officials correspond with each other reflects their relative positions in the social and administrative hierarchy, as indicated by the use of certain forms of address, honorific epithets, and request verbs.

A distinction can be made in this respect between 'explicit' and 'implicit' markers of specific social dimensions, such as social distance:[2] whereas certain linguistic forms, such as pronouns of address, are conventionalised markers of

[1] See e.g. Labov (1966).

[2] Ochs (1996: 412–3) notes that some social dimensions are conventionalised more often than others.

social distance, other forms mediate social distance in an indirect way. Languages such as Ancient Greek, which do not present a lot of these explicit markers, form an 'open' system of denoting social distance; languages such as Korean, on the other hand, form a 'closed' system (cf. McBrian 1978). For languages of the former type, the occurrence of a given feature can be considered in terms of 'probability': in certain social environments, there is a greater probability that a given feature will occur.[3]

Linguistic variation is also intimately tied to diachronic change: as Winter (1999: 68) notes, without the concept of competing, co-existing variants, and the eventual survival of just one of these variants, it would be impossible to understand the phenomenon of linguistic change. Variation in particular domains of grammar may be limited, or more extensive in nature: usually, however, when thorough-going, structural changes occur, such as the loss of the infinitive in Greek, there is a lot of synchronic variation and competing variants (cf. Leiwo 2012: 2). Similarly to synchronic variation, diachronic change can be related to probability:[4] it involves the gradual altering of the probability that a feature will occur in a certain social environment.[5]

While scholarly interest in linguistic variation and change was originally primarily oriented towards spoken language, which can be most accurately observed and analyzed, nowadays scholars recognise that a sociolinguistic theory which does not include written texts is rather limited in scope, and that the analysis of spoken language is not unproblematic either.[6] Such insights have led to the establishment of a new discipline, called 'historical sociolinguistics',[7] which is gaining ground in Classics and Byzantine studies.[8]

Scholars working on written texts have also brought to the attention that while language is of prime importance when it comes to conveying social meaning, other elements, too, are at play in written texts. Fournet (2007), for example, has argued for a 'paléographie signifiante', noting that 'l'analyse matérielle d'un document peut être porteuse de sens' (2007: 353), not only when it comes to text type, but also with regard to the socio-cultural context of writing, and the provenance of the document. Such a focus on the social significance of extra-

---

[3] For such a probabilistic approach, see e.g. Halliday (2004[1991]).

[4] See e.g. Halliday (2004[1991]: 60); Halliday and Matthiessen (2014: 73–4).

[5] In terms of expansion or restriction of occurrence.

[6] See further Bentein (2013).

[7] For an introduction, see e.g. Hernández-Campoy and Conde-Silvestre (2012); Auer *et al.* (2015).

[8] See further Bentein (2016b, 2019).

linguistic factors ties in with recent developments outside of Classics: scholars such as Bob Hodge, Gunther Kress and Theo van Leeuwen[9] have developed a framework in which they combine insights from semiotics and sociolinguistics, called 'Social Semiotics', which attempts 'to describe and understand how people produce and communicate meaning in specific social settings' (Kress and van Leeuwen 2006[1996]: 266). One of the main points of attention in this framework is the 'multi-modal' nature of communication: next to language, we also use visual, gestural, musical, choreographic, and actional resources to make meaning (cf. Lemke 1998). Pioneering multimodal research has been done in the fields of architecture, visual images, sculpture, science, and mathematics, providing insight into the nature of both *intra-semiosis* (meaning made *within* different semiotic resources) and *inter-semiosis* (meaning made *across* semiotic resources).[10]

Remarkably, however, the new discipline of Social Semiotics is currently entirely restricted to the analysis of modern-day texts. Parallel to what we have seen with the development of sociolinguistics, social semioticians show little interest in texts from the past. One recent textbook, for example, is entitled 'Multimodality. A social semiotic approach to *contemporary* communication' (Kress 2010; my emphasis). Scholars working in Classics, papyrology, and related disciplines from their side have not paid much attention to the framework either: as Fournet (2007: 353) observes when discussing his concept of 'paléographie signifiante', material and formal characteristics of texts have been of lesser interest to papyrologists, who are in the first place interested in the actual contents of the texts (understandably, of course).[11] Whereas a number of ongoing research projects have started to draw attention to material aspects of texts and their contexts of production, they have mostly done so unsystematically, on a relatively limited scale, and without integrating modern-day theoretical insights.[12]

The main goal of this paper is to introduce the concept of 'multimodality' in studies of Antiquity, papyrology in particular. My objectives are threefold: I want to outline which 'modes' can be drawn into the discussion, to analyze what sort of variation can be encountered in documentary texts for each of these modes, and to relate this variation to different types of social factors (§3). What is

---

[9] See e.g. Hodge and Kress (1988); van Leeuwen (2005); Kress and van Leeuwen (2006[1996]); Kress (2010); Hodge (2016).

[10] For this distinction, see Lim (2004: 220–221).

[11] Notwithstanding foundational studies such as Turner (1977, 1978).

[12] For some recent studies, see Yuen-Collingridge and Choat (2012); Fournet (2004; 2007); Sarri (2018).

presented here does not aim to be a full-fledged study of the topic, and is primarily exploratory in nature: as a case study, I look into one, fourth-century archive, the so-called Nepheros archive (§2), which displays interesting signs of variation in various regards. I conclude the article by briefly discussing a number of avenues for further research (§4), including the question of the complementarity between the different modes that have been analyzed (as well as the social values attached to them), and the relevance of multimodality for scribal work. Although I do not directly focus on scribes, much of what I have to say, especially concerning extra-linguistic elements such as document type and lay-out, is quite intimately tied to scribal production.

2 The Nepheros archive

Recent studies on linguistic variation in documentary texts have stressed that archives, that is, groups of texts that have been collected in antiquity by persons or institutions, for sentimental or other reasons (cf. Vandorpe 2009), represent an important source of information. Papaconstantinou (2010:13), for example, has noted that for the study of bilingualism, bilingual archives are the only source of information that is even remotely similar to what we can obtain from a modern society, as they make it possible to analyze the linguistic behavior of a restricted group of related people.[13]

In this chapter, I focus on one specific archive, the so-called 'Nepheros'-archive (IV CE), which is the last of three archives dealing with a monastic community living on the desert edge of the Herakleopolite and Upper Kynopolite nomes in Middle Egypt, more specifically the Melitian community,[14] the other two archives being the Paieous archive (mid 330s) and the Papnouthios archive (early 340s?). The archive contains 42 texts,[15] spanning a broad time period of one hundred and twenty-five years.[16] Most of the texts in the archive are letters

---

[13] For dedicated linguistic studies of archives, see e.g. Evans (2010); Vierros (2012); Bentein (2015, 2017a).

[14] The Melitians were one of several schismatic sects, named after Meletius of Lycopolis.

[15] The archive was edited by Kramer and Shelton (1987). Images of the texts in the archive can be found in this edition, and online at http://www.rzuser.uni-heidelberg.de/~gv0/Papyri/P.Neph.html. Andrea Jördens has kindly allowed me to reproduce some digital images of the texts in the archive (see Appendix).

[16] That this period is so broad can be attributed to the presence of a number of old contracts (cf. Kramer and Shelton 1987: 5). Wipszycka (2009:82) argues that the archive was created in 352 CE, or slightly afterwards.

(20 texts) or contracts (10 texts). The archive also contains two accounts, an order, a receipt, and a number of fragments.

The three above-mentioned archives provide the earliest witnesses of monasticism in Egypt, which arose in the fourth century CE. The Nepheros archive has particular value in this regard: it is not simply limited to letters coming from and going to the main figure, Nepheros, but offers a relatively complete picture of the community's dealings, as Bagnall (1993: 308) writes:

> 'Letters and contracts show an endless flow of goods in and out of the monastery, journeys by monks, prayers and requests for prayers by lay supporters, the borrowing of commodities, the buying and selling of real property, and the involvement of the clergy in the affairs of the neighboring villages'.

The main figure of the archive is, unsurprisingly, Nepheros, a priest (*presbyteros*) with a leading role in the so-called Hathor monastery, and with spiritual responsibility of the village community of *Nesoi*. In the archive, Nepheros is not just engaged with church affairs and spiritual matters. He also concerns himself with profane matters: his correspondents ask him for practical assistance in a number of matters. Paulos, for example, asks that bread should be baked in the monastery and sent to Alexandria (P.Neph.1); the monk Kapiton has lost his clothes, and Nepheros has to make sure he gets them back (P.Neph.11). Nepheros is nowhere addressed with a title, but Shelton and Kramer (1987: 9) hypothesise that he must have had a high rank in the Hathor monastery, on the basis of the broad range of activities he is involved in, and the respect he is addressed with: the editors consider him to be a προεστῶς ('prior'), perhaps even the most important προεστῶς of the monastery.

Another figure playing an important role in the archive is Paulos, who is the sender/addressee of one out of four texts in the archive (ten letters). At the time of the correspondence, Paulos lives in Alexandria, together with his wife Tapiam and their children. Since Tapiam mentions in the first letter of the archive, P.Neph.1, that she is ill, and since she is no longer referred to in the remainder of the archive, one could assume that she has died in the meantime (cf. Shelton and Kramer 1987:24). In his letters, Paulos appears as a businessman, selling the merchandise coming from the Hathor monastery by boat, and buying other things which the monastery needs. Paulos appears to be more than a mere business connection, however: he is well acquainted with the people in the Hathor monastery and its surroundings, and has a close connection with Nepheros. Shelton and Kramer (1987:25) conjecture that Nepheros may have been Paulos' former priest, responsible for the latter's spiritual education.

3 Aspects of variation in the Nepheros archive (IV CE)

Matthiessen (2007:24–25) recognises three major communicative 'modes' which are operative on the printed page: (i) written language; (ii) visual paralanguage (e.g. font family, type face, layout); and (iii) visual (pictorial) elements (e.g. drawings, paintings, photographs, maps, graphs). Since visual elements such as drawings are present only to a limited extent in documentary texts, attention will go to the first two of these modes, which can be referred to as the 'verbal mode' and the 'visual-graphological mode'. In what follows, I discuss two major aspects for each of these two modes: language choice (§3.1) and linguistic register for the verbal mode (§3.2), and document format (§3.3) and lay-out (§3.4) for the visual-graphological mode.

3.1 Language choice

Even though Egypt had a local language which was ancient and prestigious, successive invaders brought with them their own language which they used in the administration of the country. The best known example of this is of course Greek, which was introduced in Egypt with the conquests of Alexander the Great in the fourth century BCE. Although the use of Greek had first been limited to the administration and among the city élites, it became a second language for an increasing part of the population, and for some even the first language (cf. Papaconstantinou 2010: 4), through intermarriage, imitation, and the will for upward social mobility. After the Roman conquests, a Latin-speaking element was established, too, but this remained fairly marginal,[17] being restricted to administrative and military circles. Its users were normally Greek speakers, who used Latin 'as an expression of power or as a statement of adherence to the new rulers' (Papaconstantinou 2010: 5).

   As for the indigenous language, Egyptian, a new script was developed to record the contemporary vernacular, Coptic, which was based on the convenient alphabetic Greek script. During the Roman occupation, Egyptian had lost its various scripts – hieroglyphic, hieratic, and even demotic – due to the relative complexity of these respective writing systems. As such, native Egyptians were without a script: as Richter (2009: 415) writes, 'from about 100 CE until the emergence of Coptic, it was nearly impossible to correspond in the Egyptian

---

[17] Cf. Fournet (2009: 418). Fournet (2009: 421) notes that there are some 565 papyrological documents written partly or entirely in Latin.

vernacular: during a period of almost 200 years, an Egyptian native speaker not conversant with Greek had to hire a translator even to write and read letters.' In the fourth century, a new script was developed, a development which traditionally has been related to the need of Christianity and other *Buchreligionen* to reach the non-Hellenised Egyptian (village) population (cf. Bowman 1986: 158). Richter (2009) has recently drawn attention, however, to the fact that the origins of Coptic can actually be traced back into pre-Christian times. Without neglecting the role Christianity must have played, he stresses the needs of everyday written communication as a catalyst accelerating the rise of a Coptic *Schriftkultur* (cf. Richter 2009: 414–416).

In the Nepheros archive, both Coptic and Greek are attested: Greek is used in the majority of the texts, but interestingly the archive also contains two Coptic letters, P.Neph.15 (see Appendix) and P.Neph.16. The name of the sender of the first letter, P.Neph.15, has been clearly preserved: ⲁⲡⲁ ⲡⲁⲡⲛⲟⲩⲧⲉ (*Apa Papnoute* 'Father Papnoute'). That of the second letter, P.Neph.16, is less clear: we only have ⲡⲁⲡⲛ( ) ⲡⲡⲣⲉⲥⲃⲩⲧⲉⲣⲟⲥ (*Papn… Presbyteros* 'Papn… Priest'). While we could be dealing with one and the same person, the editors of the archive, Shelton and Kramer (1987), believe we are dealing with two distinct persons: linguistically, both texts are written in the Sahidic Coptic dialect, but the second letter has a light tendency for Achmimic forms.

The first of the two letters was addressed to Nepheros and a certain deacon Paieu; the second may have been addressed to Nepheros, but this is not entirely certain. Together, the two two letters show that Nepheros did not only receive his correspondence in Greek, but also in Coptic (cf. Shelton and Kramer 1987: 80). While perhaps we should keep open the possibility that Nepheros did not speak, read or write Coptic, it seems very likely that he did. As Shelton and Kramer (1987: 80) note, it would have been difficult for Nepheros to assume a role in the village community (what the editors refer to as Nepheros' 'seelsorgerischen Pflichten') without knowing Coptic.

Scholars who have studied the use of Coptic in Egypt, such as Richter (2009) and Fournet (2009),[18] have noted that its use expanded quite dramatically after the fourth century CE, at the expense of Greek. Studying the relationship between the two languages has not been without difficulty, however: as Fournet (2009: 432–5) notes, criteria such as ethnicity, religion, and social and economic status are complex and often difficult to apply. Fournet (2009) therefore proposes to relate language use to text type. From this perspective, one can say that while

---

[18] Other relevant studies include Choat (2006) and Clackson (2010).

Coptic initially was limited to letters, private accounts and religious texts, later on it expanded to other text types: from the sixth century, we see that Coptic is no longer restricted to private affairs, but is also used for legal documents, a context which before was exclusively limited to Greek (cf. also Richter 2009: 404–6).

The two Coptic letters in our archive fit well in what Fournet (2009) calls the 'first stage of Coptic usage',[19] which he dates from the fourth century to about 570 CE, and in which Coptic was mainly used for private matters. Fournet (2009: 435) notes with regard to the use of Coptic in another archive, the Paieous archive, that 'it is worth noting also that the three Coptic letters in the dossier of Paieous concern very down-to-earth subjects (making clothes, sending provisions), in contrast to some of the Greek letters, which deal with important subjects like the relations between the Melitian congregation and the bishop of Alexandria'. For the Nepheros archive, this is more difficult to evaluate, since most of the letters are relatively down to earth, with the exception of P.Neph.20, an official letter. Moreover, as Shelton and Kramer (1987) note, the contents of the first Coptic letter, P.Neph.15 is not entirely clear: the author may be trying to convince Paieu to undertake a journey to the south. P.Neph.16 is clearer: the letter involves some business matters, concerning talents. What the two Coptic letters have in common, however, is that (i) both the sender and the addressee(s) belong to a monastic milieu, and that (ii) the sender asks the addressee(s) to greet a whole range of common acquaintances (people in Nepheros' circle, that is). These two elements give us a good indication of the informal character of the correspondence, and of the low degree of social distance between the sender and addressee(s) (cf. Shelton and Kramer 1987: 34).

That the choice for Coptic was very consciously made can be seen in P.Neph.15, where the sender/author intermingles Coptic with Greek: in line five, he starts writing his greetings in Greek (ἀσπάζομαι τούς), then changes his mind, crosses the phrase out, and writes more extensive greetings in Coptic at the back side. The address itself is made in Greek, which, Bagnall (2011: 88) suggests, may have been read more easily by letter carriers.

3.2 Linguistic register

While most of the texts in the Nepheros archive are composed in Greek, they are not uniformly written in the same linguistic register. As the common or Koine language of a great number of speakers in the South-Eastern Mediterranean,

---

[19] Not including pre-Coptic, on which see Richter (2009: 406–8).

Greek underwent a quite fundamental restructuring at all linguistic levels:[20] the optative was lost, the perfect and future tense disappeared, word order changed, vocabulary was extended through derivation and borrowing, etc.[21] These changes were not uniformly adopted, however: in our written sources, a great number of registers can be found, which are situated in between linguistic innovation and archaism, with on the one hand the 'vernacular' or 'popular' Koine known from the papyri, and on the other the 'literary' Koine of Polybius and Diodorus.[22] From this point of view, the term 'Koine' can be seen as a sort of 'umbrella-term', 'a handy but unsatisfactory and idealised shorthand for several complex linguistic situations' (Swain 1996: 19).

Whereas documentary texts are typically situated towards the vernacular side of this linguistic continuum, they do not represent a homogeneous corpus, as I have argued before (cf. Bentein 2015). In order to illustrate the same point for the Nepheros archive, we can compare four letters from the archive: P.Neph.4 (see Appendix), P.Neph.19, P.Neph.12 (see Appendix), and P.Neph.20.

P.Neph.4 is a letter from Paulos to Nepheros. Paulos' language use is relatively high-level and shows some sign of education: for example, we see that Paulos consistently uses a rather wide variety of particles, including ὅθεν (l. 10), οὖν (l. 18), γάρ (l. 23), and especially δέ (ll. 16, 22, 25, 29). In another of Paulos' letters, P.Neph.9, we even find the highly classicizing particles τοίνυν (l. 7) and τοιγαροῦν (l. 11). At the same time, it must be noted that Paulos is not very consistent in his choice of higher-level features: for example, he prefers the use of ἵνα with the subjunctive after verbs of ordering such as παρακαλῶ and ἀξιῶ, rather than ὅπως with the subjunctive or the accusative and infinitive (e.g. ll. 10–11: παρακαλῶ, ἵνα μου μνημονεύσῃς).[23] Other features correspond to the Classical standard in form, but not in usage: in the relative clause ἐν οἷς δὰν βούλοιτο (ll. 5–6), for example, we find an optative after a main verb in the present tense,[24] and the particle δέ in a post-posed generalizing relative clause.[25] In terms of phraseology, some of Paulos' expressions are unusual; so, for example, the phrase γινώσκων κέρδος εἶναί μοι ψυχῆς τὸ μνημονεύεσθαι ἐν ταῖς σαῖς

---

[20] This restructuring is sometimes referred to as a 'simplification', cf. Dickey (2009: 154).

[21] See e.g. Browning (1983); Horrocks (2010); Bentein (2016a).

[22] On linguistic register, see further Bentein (2013, 2015, 2016a).

[23] On Post-classical complementation, see further Bentein (2017b).

[24] Compare Horn (1926: 144–5); Mandilaras (1973: 286).

[25] Compare Tabachovitz (1943: 26–9), followed by Youtie (1973: 116). Tabachovitz suggests to interpret δάν as a morphological reanalysis of ἐπειδάν.

εὐχαῖς τῆς σῆς εἰς θεὸν ἐλπίδος (ll. 7–10) 'knowing that it is an advantage for my soul that I am mentioned in your prayers of your hope in God': the use of both the dative and the genitive indicating advantage (μοι and ψυχῆς, to be taken with εἶναί and κέρδος respectively), the addition of the genitive τῆς σῆς εἰς θεὸν ἐλπίδος after ἐν ταῖς σαῖς εὐχαῖς, and the repetition of the possessive pronoun (σαῖς … σῆς) are all uncommon. The editors of the archive, Shelton and Kramer (1987: 20–1), have observed that Paulos' letters contain other unusual phrases, too, which are otherwise unattested, such as ἐρρῶσθαι ψυχῇ καὶ σώματι[26] (P.Neph.1, l. 30); θεοῦ δὲ ἐπινεύοντος[27] (P.Neph.6, l. 8); ὁ τῶ(ν) [ὅλων δεσπότης][28] (P.Neph.5, ll. 24–5). Since such phrases are definitely not orthodox, the editors speculate that they may be viewed as indications of the Melitian background of the correspondents in the archive.

A text of a lower linguistic level is P.Neph.19, a letter written from the village community of Neson Kome to a certain Paulos, who is addressed as τῷ κυρίῳ μου τιμιωτάτῳ πατρί, and should therefore not be identified with the author/sender of P.Neph.4 and other letters in the archive. As Shelton and Kramer (1987:89) note, contentwise this is one of the strangest documents in the archive: as far as we can tell, something has happened in the village,[29] of which the inhabitants suspect that Paulos could make a complaint to the authorities. Since such a complaint could have far-reaching consequences, the inhabitants of the village implore Paulos not to make it. The document contains several substandard features, such as the use of asyndetic parataxis after the verb παρακαλῶ (παρακαλοῦμέν σε, δέσποτα πάτερ, μὴ θελήσῃς ἐντυχῖν (ll. 4–5); παρακαλοῦμεν, ὅλον τὸ κοινὸν τῆς κώμης, μηδὲν κινήσῃς (ll. 14–16)), the use of ἑαυτῶν as the first person plural reflexive pronoun (ἡμεῖς ἀφ' ἑαυτῶν ποιοῦμεν (l. 12)) (cf. Gignac 1981: 167), the use of τό instead of ὅ as a relative pronoun (το θέλεις ποιοῦμεν (ll. 17–18)), and, perhaps most noticeably, the use of the futuristic present (cf. Mandilaras 1973: 102–5).[30] Future forms seem to be avoided altogether: forms such as γίγνεται (l. 10) 'it will happen', ποιοῦμεν

---

[26] Instead of simple ἐρρῶσθαι.

[27] Instead of θεοῦ δὲ θέλοντος.

[28] Instead of κύριος. Note that the reading ὁ τῶν ὅλων δεσπότης is not very secure. Shelton and Kramer (1987: 21) consider it 'sehr wahrscheinlich.'

[29] Shelton and Kramer (1987: 90) make a couple of suggestions: a conspiracy, a pagan ritual, or a necking party.

[30] The futuristic present can be found in other texts in the archive as well. Compare e.g. P.Neph.10, l. 15.

(l. 12) 'we will do', ἐρημώννουσιν[31] 'they will destroy'[32] are used.[33] A number of other letters in the archive can be placed in the same linguistic category, such as P.Neph.10, from the monk Horion to Nepheros, and P.Neph.11, from the monk Kapiton to two priests called Nepheros.

P.Neph.12, a letter from the monk Serapion to a certain Apa Petechon, contains the most non-standard forms. Serapion writes from Omboi, in upper Egypt, to the Hathor monastery, greeting all his brothers, including the main figure of our archive (referred to as τὸν πατέρ[α]ν μου Νεφερως τὸν πρεσβύτερον in l. 3). Serapion has serious issues with case usage: in his letter, we find various types of case interchange.[34] Quite prominent is the use of the nominative case instead of the accusative case for personal names.[35] Almost all of the personal names (mostly used after the verb ἀσπάζομαι) remain in the nominative case: so e.g. Νεφερως (l. 3); Εὐφελις (l. 5); Βησας (l. 5). This also includes personal names from the second declension in -ος, such as Συρος (l. 9), Ἑρμεινος (l. 11), Λωτός (l. 11), and Πετρος (l. 16). Sometimes complements to personal names are inflected, as in τὸν πατέρ[α]ν μου Νεφερως (l. 3); τὸν Ἰακώβ (l. 5); σὺν τῷ ἀδελφῷ αὐτοῦ Συρος (l. 9). Most often, however, this is not the case. So, for example, πατήρ is used both in the dative and in the nominative case in the first line: τῷ ἀγαπητῷ μου πατρὶ ἄπα Πετεχῶν ὁ πατὴρ τῆς μονῆς. Similarly, we read ἀσπάζομαι Βησας μοναχος (l. 5); προσαγορεύω Κασις ὁ ἀδελφος (l. 6); ἀσπάζομαι Ἑρμεινος ὁ ἀναγνωστης (ll. 10–1); γράφω σοι, ἄπα Νεφερῶς (l. 12); ζήτησον Παυλος ναυτης (l. 13). Furthermore, there is also case interchange between the genitive and dative case (so e.g. προσέχης … μου (ll. 3–4); δέξαι αὐτὰ παρ' αὐτῶ (l. 14)), between the genitive and accusative case (ἔχει … δύο λιβιτου (l. 14)), and between the dative and accusative case (π[ρ]οσαγορεύω σοι (ll. 4–5)). Shelton and Kramer (1987:74) observe that 'das Griechisch des Serapion ist desolat', and suggest that he may have been a Copt. This suggestion is confirmed by the fact that Serapion twice inserts the Coptic letter Horeh in the village name Tahmuro (Greek Ταμμωρου), which he writes in two different ways (Ταϩμουρώ (l. 11); Ταϩμ[ο]υρού (l. 17)), and that he uses the Coptic possessive

---

[31] Shelton and Kramer (1987: 91) note that P.Neph.19 contains the first attestation of this form of the verb (previously ἐρημόω), which continues to be used in Modern Greek.

[32] Shelton and Kramer (1987: 91) note that in principle ἐρημώννουσιν could also mean 'they are destroying'. This seems a rather unlikely interpretation to me.

[33] On the disappearance of the future tense, see e.g. Bentein (2014).

[34] On case interchange, see e.g. Stolk (2015).

[35] For further discussion, see e.g. Vierros (2012: 140ff., 147ff.).

prefix ⲡⲁ, rather than Greek ἀπό + genitive to refer to one person's origin (Ὧρ ⲡⲁ Ταϩμουρώ 'Hôr the one from Tahmuro'). Shelton and Kramer (1987: 74, 77) also observe that Serapion refers to Neson Kome (Νήσων Κώμη) in the letter address with the unusual εἰς Μουὴ κώμη(ν). This, they suspect, may be a Coptic rendering of the village name.[36]

A letter which stands out because of its elevated language use is P.Neph.20, an official letter from a *speculator*[37] to the *riparius*[38] of the Herakleopolite nome, Fronto, about certain payments which were due by the inhabitants of Neson Kome. While the letter is not entirely preserved (the beginning lines are missing), we immediately see a number of higher-register features. This includes the use of classicizing particles and particle combinations such as τε (l. 3), οὐ μόνον … ἀλλὰ καί (ll. 4–6), τοίνυν (l. 7), and μήτε μήν (l. 11), next to more common particles such as ἀλλά (l. 5), δέ (l. 8), and γάρ (ll. 12, 16, 17). Another noticeable characteristic is the frequent usage of future and perfect forms, such as ἑστῶτι (l. 5), πεπλήρωκεν (ll. 7–8), πεπληρῶσθαι (l. 8), προειρημένοι (l. 12), ἀρνήσονται (l. 13), and πεποιῆσθαι (l. 13). The accusative and infinitive is the preferred complementation pattern after various verbs, including δηλῶ (l. 8), ἀρνοῦμαι (l. 13), παρακαλῶ (l. 14), and εὔχομαι (ll. 19–20). The contracts in the archive, too, feature the same linguistic characteristics, but here they typically form part of formulaic phrases. Phrases such as τιμῆς τῆς συμπεφωνημένης 'for the price which has been agreed upon' (P.Neph.29, l. 11); διὰ τὸ οὕ[τ]ως συμπεπεῖσθαί με 'because I have been so convinced' (P.Neph.31, l. 9); ὁμολογῶ ἐσχηκέναι 'I agree to have received' (P.Neph.32, ll. 7–8); ἔκ τ' ἐμοῦ καὶ ἔκ τε τῶν ὑπαρχόντων [μοι πά]ντων 'against me and all my possessions' (P.Neph.32, ll. 15–6); [ἀπ] ε̣σχήκαμ[εν] παρὰ σοῦ 'we have received from you' (P.Neph.43, l. 4), containing particles, perfect/ future tense forms, and infinitival complement clauses, all form part of standard legalese in Antiquity.

One linguistic characteristic that is well attested in all of the texts in the archive, including the official letter P.Neph.20, is orthographic misrepresentations of, especially when it comes to vowels: for example, we find αι instead of ε (e.g. εὔξασθαι, P.Neph.1, l. 12), ε instead of αι (e.g. προσευχέ, P.Neph.10, l. 8; ἀσπάζομε, P.Neph.15, l. 5; κερῷ, P.Neph.18, l. 12), η instead of ε (e.g. ἤσμεν, P.Neph.10, l. 16), ι instead of ει (e.g. εὐλαβίας, P.Neph.3, l. 4; ἐκτίσῳ,

---

[36] In the Trismegistos database, ⲙⲟⲩⲉⲓ is mentioned as the Coptic name of Neson Kome (https://www.trismegistos.org/place/3045).

[37] A high-ranking official belonging to the *officium* of the prefect.

[38] The highest police official in the nome.

P.Neph.32, l. 13; χρίαν, P.Neph.4, l. 17), ει instead of ι (e.g. βεί[κους], P.Neph.29, l. 5; εὐλογείας, P.Neph.3, l. 3; ξενειτίαν, P.Neph.1, l. 10), υ instead of οι (e.g. κυτωνάριον, P.Neph.48, l. 12), ι instead of οι (e.g. ἴδατέ, P.Neph.18, l. 12), etc. Some mistakes against the consonants can be found as well: so, for example, τ instead of δ (e.g. τράπανη for δρέπανα, P.Neph.12, l. 14), δ instead of τ (e.g. μήδε, P.Neph.20, l. 11), θ instead of τ (e.g. ἐ]θελειώθη, P.Neph.29, l. 21), κ instead of γ (e.g. ἐνκεγρ[α]μμένοις, P.Neph.31, l. 17), and τ instead of ντ (e.g. π̣ετακόσια, P.Neph.33, l. 21).[39]

3.3 Document format

In the next two Sections, I discuss two visual-graphological dimensions along which variation can be found: document format and lay-out. As Sarri (2018: 87) notes, these two concepts are related and often used interchangeably. They refer to two different aspects of the text, however: 'the term format is used to refer to the shape, size and orientation of the sheet on which the letters stands, while the term layout is used to refer to the shape of the text on the sheet'. As I mentioned in the introduction, such 'material' aspects of documentary texts have not received much attention. They, too, allow for variation, however, and can therefore be considered to function as semiotic resources carrying social meaning.

Various writing materials were used for everyday writing in Antiquity: papyrus is most often attested, but other writing materials were also used, including potsherd, wood, linen, leather and parchment (cf. Bülow-Jacobsen 2009). Variation in this regard is not attested in the Nepheros archive: all of the texts in the archive are written on papyrus. Interestingly, however, our texts do not have the same format. Fournet (2007), based on earlier research by Turner (1978), has proposed to distinguish between two main types of document format, which he calls the 'vertical' and the 'horizontal' format.[40] As figure 1 shows, these two types of format have quite different characteristics: the vertical format is high

---

[39] Note that the phonological changes indicated by these orthographic mistakes represent different phases in the history of Greek. So, for example, the interchange between ι and ει is common already in Early Ptolemaic times, whereas that between υ and οι and that between ι and οι becomes more common only in the Roman period. For further discussion, see Gignac (1976); Teodorsson (1977); Horocks (2010). For a comprehensive overview of interchanges, see the *Trismegistos Text Irregularities Database* at https://www.trismegistos.org/textirregularities/index.php.

[40] Sarri (2018) uses the terms '*transversa charta* format' and 'pagina format'. Sarri (2018: 95–7) recognises a third type of format, the 'Demotic style format', which, she argues, was abandoned by the second century BCE (Sarri 2018: 97).

but not broad, and has the writing along the fibers, allowing for relatively short lines. The horizontal format on the other hand, is broad but not high, and has the writing across the fibers, thus allowing for very long lines.[41]
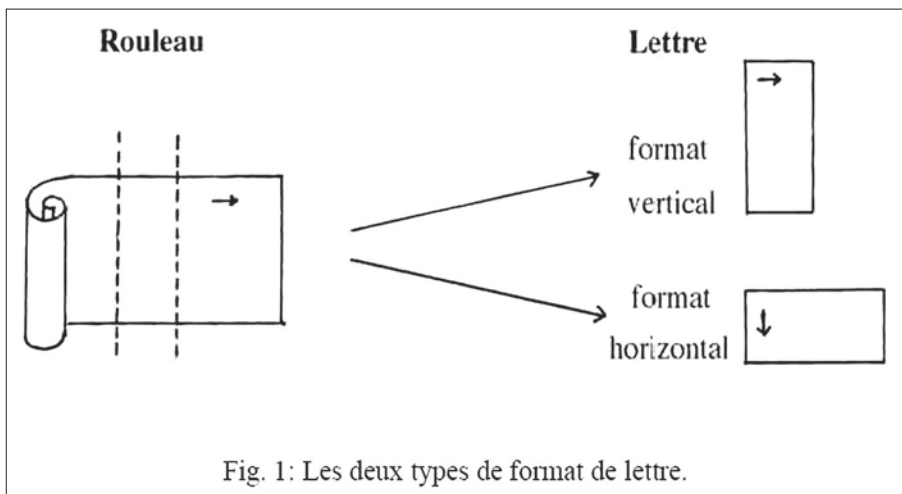
Figure 1 (from Fournet 2007: 354).

While some editions explicitly indicate whether the writing is along or across the fibers, this practice is not uniformly adopted, so that inspection of the text is needed. One element that clearly shows the difference between the two types is the *kollêseis* or glue joins: with the horizontal format, writing and *kollêseis* run in the same direction. Another element is standard size: Turner (1978: 15, 61) notes that the height of an average *kollêma* or constituent sheet was around 28 to 30 cm,[42] and that its breadth/width did not exceed 20 cm. If a text is much larger than this, we can expect *kollêseis*.[43] Starting from the fourth century, however, one begins to find sheets of greater bredth, which Turner (1978: 62) interprets as a change in the manufacturing process. Finally, having a look at the verso side of the text also helps: here one can see the orientation of the fibers more clearly.

Fournet (2007) has drawn attention to a conspicuous relationship between text type and document format: in Late Antiquity, up to 87% of the petitions (V–

---

[41] As Sarri (2018: 91) notes, 'writing against the fibers of the papyrus was not as convenient as writing along the fibers.'

[42] It could range up to 37 cm, however (Turner 1978: 15).

[43] Compare Turner (1978: 55).

VII CE) employ the vertical format, whereas notarial documents predominantly use the horizontal format. Fournet (2004: 73) therefore suggests that document format may have been socially meaningful, that is, it may have been a way of indicating the difference between texts that were meant for the authorities (petitions) versus texts that came about between private parties by the mediation of a notary (contracts).

Both the vertical and the horizontal format are attested in our archive. The letters are generally written in the vertical format, with a height between 22 and 24 cm, and a breadth between 6 and 15 cm (standard size, that is).[44] P.Neph.16, one of our Coptic letters, seems to follow the same format, with a height of 23 cm. and a breadth of 8 cm. There are a couple of noticeable exceptions, however. P.Neph.12 (see Appendix), Serapion's letter, does not follow conventional size: as Shelton and Kramer (1987: 73) note, 'der Brief des Serapion unterscheidet sich bereits durch das Format von den anderen Briefen des Archivs'. The letter is broader than it is high, with a breadth of 27.8 cm and a height of 24.7 cm. The text seems to be written along the fibers, on a very broad papyrus sheet. The same is true for P.Neph.20 (see Appendix), our official letter, which is slightly broader than it is high, with a height of 21.5 cm and a breadth of 22.5 cm; this text, too, was written on a broad papyrus sheet. Finally, P.Neph.15 (see Appendix), the second Coptic letter, has a different format: it is horizontally positioned, and written against the fibers on the recto side.[45] It has a height of 6 cm and a breadth of 11.5 cm.

Interestingly, the contracts are not uniformly composed in the horizontal format, as one would expect: P.Neph.29 (a contract of sale) and P.Neph.31 (a cession of land) are composed in the horizontal format, but P.Neph.32 (a contract of loan) (see Appendix) and P.Neph.33 (a contract of sale) are not. One way to explain this variation is to connect it to the style in which the contracts are formulated. Contracts in antiquity could be 'objectively' or 'subjectively' styled:[46] with the former type, the point of view of an objectively witnessing third party is adopted, whereas with the latter type the perspective of the contracting parties themselves is maintained. Consequently, third person references dominate in objectively styled contracts, whereas first- and second-person references dominate in subjectively styled contracts. The above-mentioned contracts are clearly written in different styles: In P.Neph.32, for example, we find first- and second-person references such as ὁμολογῶ ἐσχηκέναι παρὰ σοῦ (ll. 7–8), ἐκτίσω σοι

---

[44] See e.g. Paul's letters, P.Neph.1, 3,4, 5, 6, 8, 10.

[45] Along the fibers on the verso side.

[46] See e.g. Mitteis (1963).

(l. 13), τῆς πράξεως σοὶ οὔσης [ἢ τοῖ]ς παρὰ σ[ο]ῦ ἔκ τ᾽ ἐμοῦ καὶ ἔκ τε τῶν ὑπαρχόντων [μοι] (ll. 14–6). In P.Neph.33, we similarly find references such as ἐλθό]ντα εἰς ἐμὲ (l. 5), ἐγὼ ὁ ἀπ[οδόμενος (l. 10), ἀπὸ τῶν ἐμῶν ἀναλωμάτων (l. 16), etc. In P.Neph.29, on the other hand, we find third-person references such as καθ[ὼς αὐτοὶ ὑ]πηγόρε[υσαν (l. 7) and τοῦ πωλο]ῦντ[ος … τοῦ ὠν]ουμένου (l. 10). Since contracts in the subjective style are much closer to letters, it is not entirely surprising that the vertical format is used. P.Neph.31 forms an exception to this general observation, since, here, too, first person references can be found: so, for example, ἐμοῦ τῆς χα[ρι]ζομένης (l. 4), των ἄλλων υἱῶν μου (l. 5), διὰ τὸ ἐμὲ εἶναι κεχαρισμ[ε]να (l. 7), etc. Since this contract is of a somewhat different nature than P.Neph.29, P.Neph.32 and P.Neph.33, all of which are contracts of loan/sale, it may be that the inherently more elevated nature of P.Neph.31, a cession of land, automatically triggered the use of the horizontal format.[47]

The largest document in our archive is not a contract, but an account written in two columns, P.Neph.38. This text has a height of 24 cm and a breadth of 33 cm.

3.4 Lay-out

To conclude, let me briefly turn to the lay-out of the texts in the Nepheros archive. As mentioned above, lay-out very broadly has to do with the shape of the text on the sheet. Recent studies have analyzed lay-out under the heading of 'typography', which, although sometimes associated specifically with printed text, is now increasingly being used to refer to the visual organization of written language in whatever way it is produced (cf. Walker 2001: 2).

Contrary to linguistics, typography does not have a formal and established descriptive tradition (cf. Walker 2001: 17, 23). As such, which typographical aspects to include remains a point of debate. In multimodal studies, a distinction is commonly made between two main levels, which are called the 'micro'- and the 'macro'-level. Whereas the micro-level has to do with the shape and size of individual letters, ligatures, and the spacing between characters and words (cf. Sijpesteijn 2007: 515), the macro-level has to do with the number and size of lines, margins, indentation, etc. Both aspects have an important social role to play, as Sijpesteijn notes:

'The shape and form of letters used to write a text affect not just its appearance but also its "atmosphere", conveying important messages about the social and

---

[47] Further research is needed, though.

professional intentions and background of the scribe and reader beyond the content of the text itself.'[48] (Sijpesteijn 2007: 515)

The micro-level, that is, the shape and size of letters, is typically treated in papyrological studies by referring to three major script types or 'hands', that is, the 'book hand', the 'chancery hand', and the 'cursive hand'.[49] In terms of this threefold distinction, we can say that all of our documents have been written in a cursive hand, which was the handwriting used for everyday business matters and daily life. Even so, we can see some noticeable differences. Bagnall and Cribiore (2006: 42–5) propose to characterise such differences by further distinguishing between three types of (cursive) hands, called 'documentary' hands, which are not particularly concerned with legibility and show rapid, ligatured cursives, 'secretarial' hands, which are characterised by legible, well-spaced handwriting, with larger letters written at a slower pace, and ligatures kept at a minimum, and 'personal' hands, which show lack of expertise in writing, and have an unprofessional air (as indicated by unruly right margins, varying line spaces, wavering lines, clumsy corrections, and retracing of letters).[50] As can be seen, these different types of handwriting can be related to a number of social factors: legibility is connected to the education and training of the person writing, to the purpose of the text, but also to the degree of social distance between the sender and the addressee.[51]

One part of the archive where interesting signs of typographical variation can be found are Paulos' letters.[52] We can contrast, in this regard, the writing style of P.Neph.1 (Paulos and Tapiam to Nepheros, Ophellios and others) and

---

[48] Sijpesteijn's (2007) observation was made with regard to Arabic papyri, but is relevant for Greek documentary texts as well.

[49] See e.g. Thompson (1912); Montevecchi (1988: 47–49). Note that the distinction between 'chancery hand' and 'cursive hand' is not absolute, since the former also contains ligatures.

[50] Obviously, these three categories form a continuum, from very experienced to very inexperienced hands. Compare e.g. Bagnall and Cribiore (2006: 45), when they note with regard to personal hands that 'at the highest end of the spectrum there are hands only slightly less practiced than some of the secretarial ones, while at the lowest end there are those hands that can be defined, in the terms used for school exercises, as "evolving", and occasionally even "alphabetic"'.

[51] Compare Bataille (1954: 77–8), who distinguishes between four different 'modes' of handwriting: 'impersonal', 'respectful', 'familiar' and 'private'.

[52] As Shelton and Kramer (1987: 24) note, Paulos' letters are written in different hands, indicating that they do not all represent autographs. It is impossible to determine, however, if there are any autographs among the letters, and if so, which hand is Paulos' own.

P.Neph.4 (Paulos to Nepheros) (see Appendix), with that of P.Neph. 3 (Paulos to Nepheros) and P.Neph.6 (Paulos to Nepheros). In P.Neph.1 and P.Neph.4, the writing is condensed, and irregular in that letters do not stay within the line, and vary in size.[53] In P.Neph.3 and P.Neph.6, on the other hand, the writing has a much more spacious outlook, and letters have a more regular appearance. Both of these texts have delicately formed letters. The letters ε and κ, especially in P.Nep.6, are quite distinctive, giving the text an angular appearance. In P.Neph.4, on the other hand, letters are much more heavily formed. Some of Paulos' other letters are in between these two extremes, such as P.Neph.5, P.Neph.8, P.Neph.9, and P.Neph.10. The reasons behind the typographical variation in Paulos' letters remain unclear. There does not seem to be a clear connection with the contents of the letters: it is true that P.Neph.1 has a more personal tone than the other letters, but P.Neph.4 does not substantially differ from P.Neph.3 and P.Neph.6 in character (all three are business letters with some personal remarks).

Shelton and Kramer (1987: 85) contrast the style of the personal letters written by men in the archive with that of P.Neph.18, a letter written by a(n Egyptian) woman called Tauak to Eudaimon and his wife Apia about an *aroura* of land.[54] They note that the writing is small and unattractive, and cannot be compared with the letters of the male correspondents in the archive. Even though the editors explicitly postulate a gender-difference here, they have to admit that it is impossible to tell whether the handwriting was Tauak's own. Regrettably, the image provided by Shelton and Kramer (1987) is rather blurry, so that a detailed analysis of the handwriting is difficult.

Two other documents which stick out from a typographical point of view are the Coptic letters in the archive: in these two texts, unligatured capitals can be found. In both P.Neph.15 (see Appendix) and P.Neph.16, upright, heavy letters are used, that have a very regular and spacious appearance. As mentioned earlier, the writer of P.Neph.15 at one point switches from Coptic to Greek, but then changes his mind. The difference in writing between the Greek and Coptic letters is quite noticeable:[55] the Greek phrase ἀσπάζομαι τούς is written in much smaller and sloping letters than the Coptic in the same text.

---

[53] For these and other descriptive dimensions, see van Leeuwen (2006), who proposes to distinguish between *weight*, *expansion*, *slope*, *curvature*, *connectivity*, *orientation*, and *regularity*.

[54] As Shelton and Kramer (1987: 85) note, the inclusion of this text in the Nepheros archive is problematic: it seems quite sure that Eudaimon and Apia did not belong to the Hathor monastery. Tauak might have, but in that case, her letter was either never sent, or sent back.

[55] As also noted by Bagnall (2011: 88).

As can be expected, all of the contracts have a professionally written appearance, as characterised by a rapid, cursive style of writing and relatively small letters. The typographical analysis of the contracts is rendered more difficult by the fact that none of them have been preserved in their entirety. Of the four earlier mentioned contracts (P.Neph.29, 31, 32, 33), P.Neph.32 (see Appendix) probably has the most distinctive appearance: its letterforms are heavier, more sloping and ligatured than in the three other contracts.[56] In all of the contracts, we also see a second hand, and sometimes even a third or fourth hand. This change of handwriting is most noticeable in P.Neph.29, where we shift in l. 23 from a professional hand first to the hand of the seller, Aurelius NN, and afterwards to the hand of perhaps a witness, Atisis.[57] Both these hands employ very large, unligatured and irregular letters, and are therefore good examples of what Bagnall and Cribiore (2006) call 'personal hands', that is, hands of individuals with much less training and experience in writing.

Two other texts with a professional outlook are P.Neph.12 and P.Neph.20. Both of these texts seem to have been written rather rapidly, in a cursive hand with quite a few ligatures (especially P.Neph.12). An interesting similarity between both texts is that the writer fills up the right margin by extending a horizontal line from the last letter. This must be related to the fact that unusually broad sheets have been used: the phenomenon is nowhere else attested in the archive.[58] As Shelton and Kramer (1987: 74) note, letters are formed quite irregularly in P.Neph.12.

One macro-level characteristic that is quite noticeable in various texts in the archive is vertical writing in the margins. When writers had reached the bottom of a column, rather than adding another column, they continued writing vertically in the margins. In most cases this was the left margin, since this was usually wider than the right one. Sarri (2018: 112) notes that this represents a rather marginal phenomenon: in Egypt, it is attested only in letters from the Roman imperial period (I – IV CE), mostly in texts from the second to the fourth century CE. As Sarri (2018: 113) writes, there seems to be a connection between writing in the margins and formality, since writing vertically in the margins is attested only

---

[56] This distinctive handwriting could, perhaps, be connected to the earlier-mentioned vertical format and the use of subjective references, but more research is needed to firmly establish such a correlation.

[57] Cf. Shelton and Kramer (1987: 108) on the person and function of Atisis.

[58] A couple of strokes are present in P.Neph.1, but these have not been drawn systematically.

in private letters, not in official ones.[59] The phenomenon is attested in six letters in the archive, four of which from Paulos to Nepheros (P.Neph.4, P.Neph.6, P.Neph.7 and P.Neph.8). It can also be found in P.Neph.17, a fragmentary letter from Apa Elo(?) to a certain Petesans, and P.Neph.18, the earlier mentioned letter from Tauak to Eudaimon and Apia. In these last two letters, writing in the margins is used for a health wish and greetings. In Paulos' letters, marginal writing is also used for more constitutive aspects: it is used in P.Neph.6, 7 and 8 for a post script. In P.Neph.7, part of the last sentence of the main body continues in the margin.

4. Concluding remarks

Whereas scholars such as Labov initially showed a negative attitude towards the sociolinguistic study of texts from the past, nowadays the analysis of linguistic variation and its correlation with social factors has become common practice. In this contribution, I have attempted to show that such linguistic variation exists in the fourth-century Nepheros archive, too. The texts in the archive can be placed on a linguistic continuum ranging from 'popular' or 'vernacular' on the one hand (e.g. case alternation, asyndetic parataxis, futuristic present) to 'classicizing' or 'archaizing' on the other hand (e.g. classical particles, infinitival complement structures).

The main point of this contribution, however, has been that variation should be conceived of as relevant to other semiotic domains as well, and not strictly limited to linguistic register: in the area of language choice, for example, we see that Coptic is used next to Greek. When it comes to document format, not one, but two formats are attested: the 'vertical' and the 'horizontal'. Most complex,[60] however, is lay-out or typography, which goes much further than a binary choice: both the micro-level and the macro-level show various types of variation. Throughout my contribution, I have drawn attention to the social factors that are related to these different types of variation, such as *social distance* (language choice, linguistic register, lay-out), *social status* (language choice), *formality* (language choice, linguistic register, lay-out), *text type* (language choice, linguistic

---

[59] As Sarri (2018: 113) notes, 'the phenomenon is also an indication that the writers wrote the letters directly, without previous drafts, which would have enabled a better estimate of the length of the sheets to be cut.'

[60] On the complexity of different semiotic systems, see e.g. Eggins (2004: 11–7).

register, document format), *education* (linguistic register, lay-out), *ethnicity* (language choice, linguistic register), *religion* (linguistic register), and *gender* (lay-out). As recent sociolinguistic research has argued, the relationship between variants on the one hand and social meaning on the other is best thought of as dynamic, rather than static and fixed: Eckert (2008), for example, has suggested that (linguistic) variants have an abstract 'indexical potential'. In her view, social values are grouped as 'indexical fields', that is, constellations of meanings that are ideologically linked.

I certainly do not want to claim that such variation has been completely ignored by previous studies (quite to the contrary). However, it has mostly been studied in isolation: many papyrologists tend not to focus on material aspects of writing, those who do, do not focus on linguistic aspects, etc. In this respect, I have suggested that the framework of 'multimodality' may provide a useful starting point for viewing our texts as the complex, multi-layered semiotic objects they are. The framework forces us to look at variation from a holistic point of view, and to look at the different elements that are involved as dimensions of 'meaning-making', as semiotic resources. An important distinction in this regard is that between 'intra-semiosis' and 'inter-semiosis', that is, meaning making *within* semiotic resources, and meaning making *across* semiotic resources.

Inter-semiosis, or, in other words, the correlation between different semiotic resources, is usually studied at some level of abstraction (for example through the concepts of register and genre), on the basis of an extensive corpus of texts. At this stage of the research, I am unable to present such high-level abstractions. However, a number of interesting correlations have become visible in the individual texts in our small corpus: P.Neph.12, for example, is poorly written (especially from the point of view of orthography and case marking), has an unusual size, and shows a rapid, cursive and ligatured, handwriting. Similarly, in P.Neph.15 and 16, the choice for Coptic, rather than Greek, correlates with a distinctive handwriting (unligatured, heavy capitals) and a distinctive document format.[61] Next to such convergences,[62] we also see some divergent features: for example, while Paulos attempts to employ higher-register Greek, the handwriting is not always very legible, and marginal writing often occurs. The same can be noted for P.Neph.20, which employs high-register Greek, but an unusual document format, and rather rapid writing. This may be taken as the combined effect of social factors which do

---

[61] Compare also Shelton and Kramer (1987: 85), when they note that the sloppy handwriting of P.Neph.18 corresponds with its use of vernacular language.

[62] What Royce (2007) refers to as 'intersemiotic complementarity'.

not come together very often in documentary texts, such as high social status and education on the one hand, and informality and low social distance on the other, but this remains to be confirmed.

Even though this volume is thematically centered around the 'acts of the scribe', I have not paid a great deal of attention to scribes.[63] Thanks to the efforts of Martti Leiwo and others, scribes have come under renewed attention: scholars have drawn attention to the fact that scribes, both professional and non-professional ones, were quite often employed in Ancient Egypt, and thus may have had an impact on the language of the texts we are studying.[64] As Bergs (2015: 117) notes, this intermediary role of scribes presents a complication for linguistic analysis: 'if we want to contextualise the producers of language and talk about the influence of external factors such as education, gender, social strata, social networks, place of living, we need to know whose language we are looking at.'[65] Evidently, this 'scribal issue' is just as much a complicating factor for other types of variation: for elements such as language choice, document format, and lay-out, too, one would like to know what kind of impact the scribe had.

I do not see scribal influence as a 'fundamental problem', though, as others do.[66] In principle, there is little reason why scribes would have been insensitive to the social context in which the document was produced. In fact, Bergs (2015: 130) has observed with regard to scribal influence in Early Modern English that scribes employed different constructions to represent their different authors, even in cases when they were not taking down dictation *verbatim*. From a multi-modal perspective, however, a whole new set of questions deserves to be answered: did scribes have more influence on one semiotic resource than on another? Were scribes aware of convergences between different semiotic resources? Were there significant differences between scribes in terms of their 'communicative competence', that is, their knowledge of variation and how to employ it? These and other questions await further research.

---

[63] Document type and lay-out are, of course, intimately tied to the work of the scribe.

[64] For a recent collection of papers on this topic, see Cromwell and Grossman (2018).

[65] Compare Halla-aho (2018) for Latin and Greek documentary texts.

[66] See e.g. Elspass (2012: 158).

Appendix
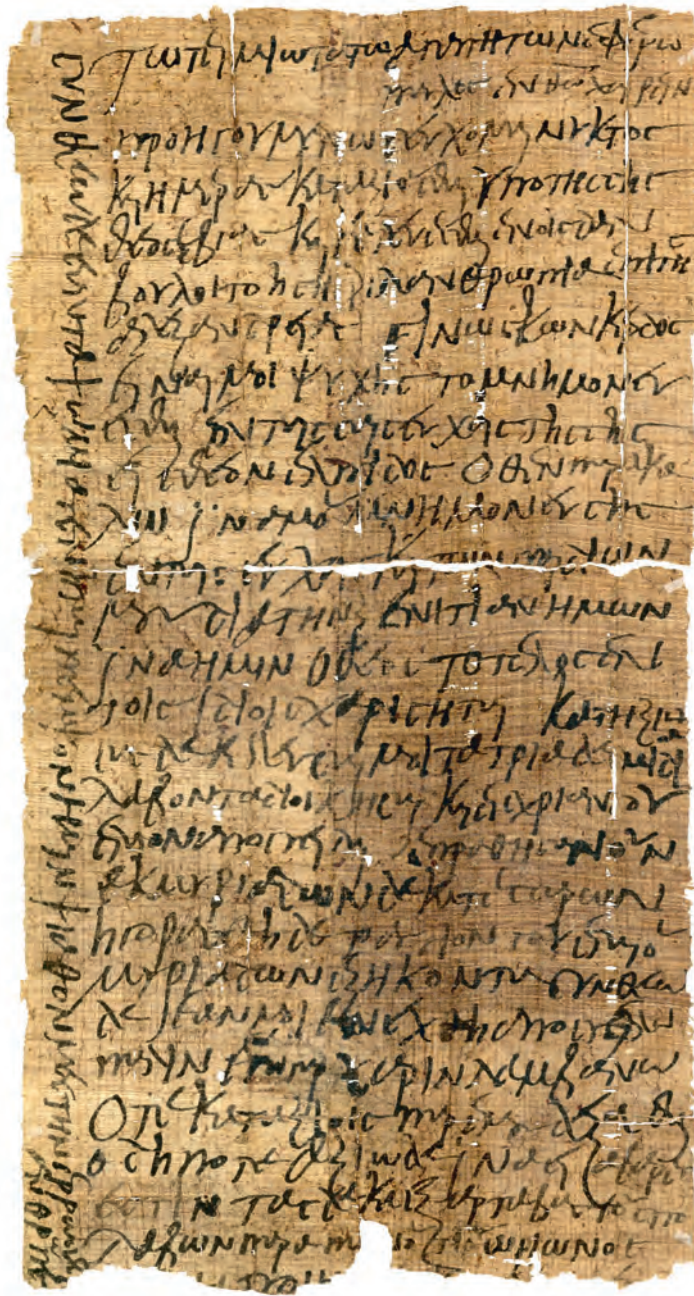
Figure 2. P.Neph.4.
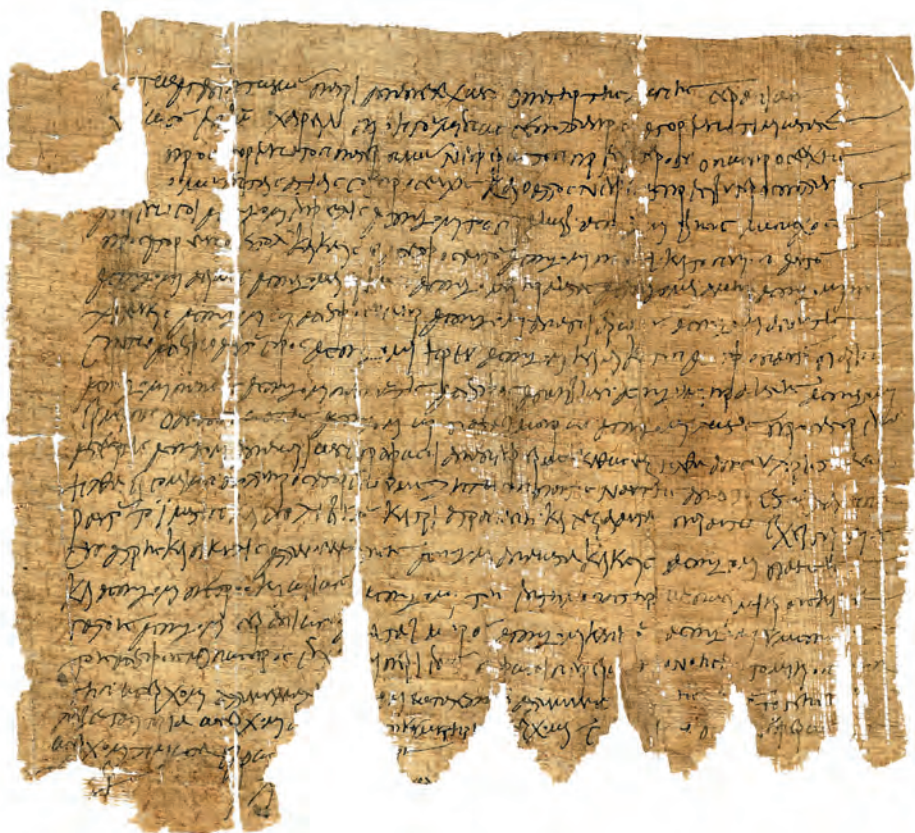
Figure 3. P.Neph.12.
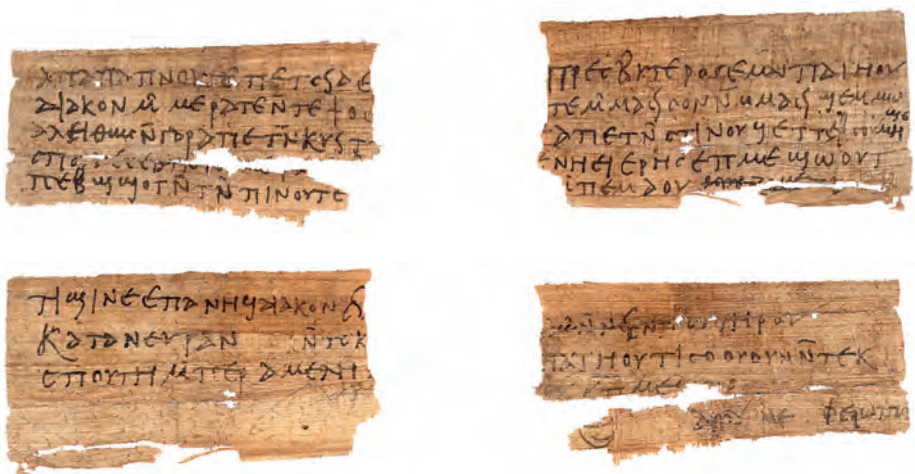
Figures 4A + 4B. P.Neph.15R.
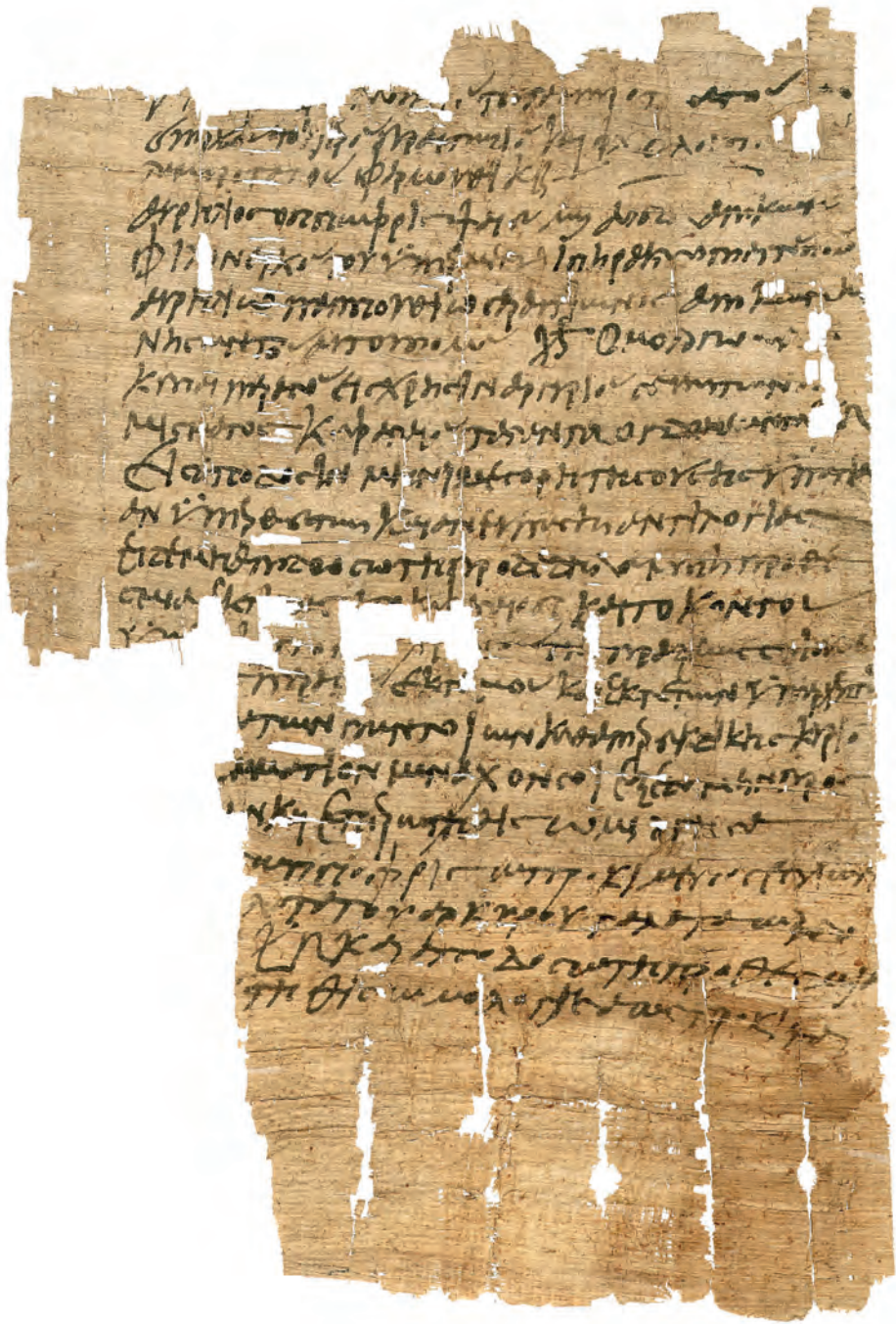
Figure 5. P.Neph.32. © Institut für Papyrologie, Universität Heidelberg.

References

Auer, Anita, Peersman, Catharina, Pickl, Simon, Rutten, Gijsbert and Rik Vosters. 2015. 'Historical sociolinguistics: the field and its future', *Journal of Historical Sociolinguistics* 1: 1–12.

Bagnall, Roger S. 1993. *Egypt in Late Antiquity*. Princeton: Princeton University Press.

Bagnall, Roger S. 2011. *Everyday writing in the Graeco-Roman East*. Berkeley: University of California Press.

Bagnall, Roger S. and Raffaella Cribiore. 2006. *Women's letters from Ancient Egypt, 300 BCE – CE 800*. Ann Arbor: The University of Michigan Press.

Bataille, André. 1954. *Pour une terminologie en paléographie grecque*. Paris: C. Klincksieck.

Bentein, Klaas. 2013. 'Register and the diachrony of Post-classical and Early Byzantine Greek', *Revue Belge de Philologie et d'Histoire* 91: 5–44.

Bentein, Klaas. 2014. 'Tense and aspect from Hellenistic to Early Byzantine Greek', in Georgios K. Giannakis *et al.* (eds.), *Encyclopedia of Ancient Greek language and linguistics*. Leiden and Boston: Brill, pp. 379–82.

Bentein, Klaas. 2015. 'The Greek documentary papyri as a linguistically heterogeneous corpus: The case of the katochoi of the Sarapeion-archive', *Classical World* 108: 461–84.

Bentein, Klaas. 2016a. *Verbal periphrasis in Ancient Greek: HAVE- and BE-constructions*. Oxford: Oxford University Press.

Bentein, Klaas. 2016b. 'Towards a socio-historical analysis of Ancient Greek? Some problems and prospects', in Andrea M. Cuomo and Erich Trapp (eds.), *Toward a historical sociolinguistic poetics of Medieval Greek*. Turnhout: Brepols, 35–44.

Bentein, Klaas. 2017a. 'διά as a polysemous preposition in Early Byzantine Greek. "Dead ends" and other uses in the Qurrah archive (VIII AD)', *Symbolae Osloenses* 91: 1–25.

Bentein, Klaas. 2017b. 'Finite versus non-finite complementation in Post-classical and Byzantine Greek. Towards a pragmatic restructuring of the complementation system?', *Journal of Greek Linguistics* 17: 3–36.

Bentein, Klaas. 2019. 'Historical sociolinguistics: how and why? Some observations from Greek documentary papyri', *Aion* 41: 145–54.

Bergs, Alexander. 2015. 'Linguistic fingerprints of authors and scribes', in Anita Auer, Daniel Schreier and Richard J. Watts (eds.), *Letter writing and language change*. Cambridge: Cambridge University Press, pp. 114–32.

Bowman, Alan K. 1986. *Egypt after the pharaohs, 332 BCE – CE 642: from Alexander to the Arab conquest.* London: British Museum.

Browning, Robert. 1983². *Medieval and Modern Greek.* Cambridge: Hutchinson.

Bülow-Jacobsen, Adam. 2009. 'Writing materials in the ancient world', in Roger S. Bagnall (ed.), *The Oxford handbook of papyrology.* Oxford and New York: Oxford University Press, 3–29.

Choat, Malcolm. 2006. *Belief and cult in fourth-century papyri.* Turnhout: Brepols.

Clackson, Sarah and Arietta Papaconstantinou. 2010. 'Coptic or Greek? Bilingualism in the papyri,' in Arietta Papaconstantinou (ed.), *The multilingual experience in Egypt from the Ptolemies to the Abbasids.* Farnham: Ashgate, 73–104.

Cromwell, Jennifer and Eitan Grossmann. 2018. *Scribal repertoires in Egypt from the New Kingdom to the Early Islamic period.* Oxford: Oxford University Press.

Dickey, Eleanor. 2009. 'The Greek and Latin Languages in the Papyri' in Roger S. Bagnall (ed.), *The Oxford handbook of papyrology.* Oxford: Oxford University Press, 149–69.

Eckert, Penelope. 2008. 'Variation and the indexical field', *Journal of Sociolinguistics* 12: 453–76.

Eggins, Suzanne. 2004². *An introduction to Systemic Functional Linguistics.* New York: Continuum.

Elspass, Stephan. 2012. 'The use of private letters and diaries in sociolinguistic investigation', in Juan M. Hernández-Campoy and Juan C. Conde-Silvestre (eds.), *The Handbook of Historical Sociolinguistics.* Chichester and Malden: Wiley-Blackwell, pp. 156–69.

Evans, Trevor V. 2010. 'Identifying the language of the individual in the Zenon Archive', in Trevor V. Evans and Dirk Obbink (eds.), *The Language of the Papyri.* Oxford: Oxford University Press, pp. 51–70.

Fournet, Jean-Luc. 2004. 'Entre document et littérature: La pétition dans l'Antiquité tardive', in Denis Feissel and Jean Gascou (eds.), *La pétition à Byzance.* Paris: Association des amis du Centre d'histoire et civilisation de Byzance, 61–74.

Fournet, Jean-Luc. 2007. 'Disposition et réalisation graphique des lettres et des pétitions proto-byzantines: Pour une paléographie 'signifiante' des papyrus documentaires', in Jaakko Frösén, Tiina Purola and Erja Salmenkivi (eds.), *Proceedings of the 24th International Congress of Papyrology.* Helsinki: Societas Scientarum Fennica, pp. 353–67.

Fournet, Jean-Luc. 2009. 'The multilingual environment of Late Antique Egypt: Greek, Latin, Coptic and Persian documentation', in Roger S. Bagnall (ed.), *The Oxford handbook of papyrology.* Oxford: Oxford University Press, pp. 418–51.

Gignac, Francis T. 1976–81. *A grammar of the Greek papyri of the Roman and Byzantine periods*. Milano: Istituto editoriale cisalpino-La goliardica.

Halla-aho, Hilla. 2018. 'Scribes in private letter writing: linguistic perspectives', in Jennifer Cromwell and Eitan Grossmann (eds.), *Scribal repertoires in Egypt from the New Kingdom to the Early Islamic period*. Oxford: Oxford University Press, pp. 227–39.

Halliday, M.A.K. 2004[1991]. 'Towards probabilistic interpretations', in Jonathan J. Webster (ed.), *Computational and quantitative studies*, 42–62. London and New York: Continuum.

Halliday, M.A.K. and C.M.I.M. Matthiessen. 2014. *Halliday's introduction to Functional Grammar*. London and New York: Taylor and Francis.

Hernández-Campoy, Juan M. and Juan C. Conde-Silvestre (eds.), *The handbook of historical sociolinguistics*. Chichester and Malden: Wiley-Blackwell, pp. 156–69.

Hodge, Robert and Gunther Kress. 1988. *Social semiotics*. Cambridge: Polity Press.

Hodge, Robert. 2016. *Social semiotics for a complex world: analysing language and social meaning*. Cambridge and Malden: Polity Press.

Horn, Robert. C. 1926. *The use of the subjunctive and optative moods in the non-literary papyri*. Philadelphia: University of Pennsylvania (diss.).

Horrocks, Geoffrey C. 2010². *Greek: A history of the language and its speakers*. Chichester and Malden: Wiley-Blackwell.

Kramer, Bärbel and John C. Shelton. 1987. *Das Archiv des Nepheros und verwandte Texte*. Mainz am Rhein: P. von Zabern.

Kress, Gunther R. and Theo van Leeuwen. 2006[1996]. *Reading images: the grammar of visual design*. 2nd edition. London and New York: Routledge.

Kress, Gunther. 2010. *Multimodality: A social semiotic approach to contemporary communication*. London and New York: Routledge.

Krüse, Thomas. 2002. *Der königliche Schreiber und die Gauverwaltung: Untersuchungen zur Verwaltungsgeschichte Ägyptens in der Zeit von Augustus bis Philippus Arabs (30 v. Chr.–245 n. Chr.)*. München: K.G. Saur.

Labov, William. 1966. *The social stratification of English in New York City*. Washington, DC: Center for Applied Linguistics.

Leiwo, Martti. 2012. 'Introduction: variation with multiple faces', in Martti Leiwo, Hilla Halla-aho and Marja Vierros (eds.), *Variation and change in Greek and Latin*. Helsinki: Suomen Ateenan-instituutin säätiö, 1–12.

Lemke, Jay L. 1998. 'Multiplying meaning: Visual and verbal semiotics in scientific text', in J.R. Martin and Robert Veel (eds.), *Reading science: critical*

and functional perspectives on discourses of science. London and New York: Routledge, pp. 87–113.

Leeuwen, Theo van. 2005. *Introducing social semiotics*. London and New York: Routledge.

Leeuwen, Theo van. 2006. 'Towards a semiotics of typography', *Information Design Journal* 14: 139–55.

Lim, Victor F. 2004. 'Developing an integrative multi-semiotic model', in Kay L. O'Halloran (ed.), *Multimodal discourse analysis: Systemic Functional perspectives*. London and New York: Continuum, pp. 220–46.

Ljungvik, Herman. 1932. *Beitrage zur Syntax der spätgriechischen Volkssprache*. Uppsala: Almqvist and Wiksell.

Mandilaras, Basil G. 1973. *The verb in the Greek non-literary papyri*. Athens: Ministry of Culture and Sciences.

Matthiessen, C.M.I.M. 2007. 'The multimodal page', in Terry D. Royce and Wendy L. Bowcher (eds.), *New directions in the analysis of multimodal discourse*. Hillsdale, NJ: Taylor and Francis, pp. 1–62.

McBrian, Charles D. 1978. 'Language and social stratification: the case of a Confucian society', *Anthropological Linguistics* 20: 320–6.

Mitteis, Ludwig and Ulrich Wilcken. 1963. *Grundzüge und Chrestomathie der Papyruskunde. Zweiter Band: juristischer teil. Erste hälfte: Grundzüge (von L. Mitteis)*. Leipzig and Berlin: B.G. Teubner.

Montevecchi, Orsolina. 1988. *La papirologia*. Milano: Vita e Pensiero.

Ochs, Elinor. 1996. 'Linguistic resources for socializing humanity', in John J. Gumperz and Stephen C. Levinson (eds.), *Rethinking linguistic relativity*, Cambridge: Cambridge University Press, pp. 407–38.

Papaconstantinou, Arietta. 2010. 'Introduction', in Arietta Papaconstantinou (ed.), *The multilingual experience in Egypt, from the Ptolemies to the Abbasids*. Burlington, VT: Ashgate, pp. 1–16.

Richter, Tonio S. 2009. 'Greek, Coptic and the "Language of the Hijra": The rise and decline of the Coptic language in late antique and medieval Egypt', in Hannah M. Cotton *et al.* (eds.), *From Hellenism to Islam: cultural and linguistic change in the Roman Near East*. Cambridge: Cambridge University Press, pp. 401–46.

Royce, Terry D. 2007. 'Intersemiotic complementarity: A framework for multimodal discourse analysis', in Terry D. Royce and Wendy L. Bowcher (eds.), *New directions in the analysis of multimodal discourse*. Hillsdale, NJ: Taylor and Francis, pp. 63–109.

Sarri, Antonia. 2018. *Material aspects of letter writing in the Graeco-Roman World. c. 500 BCE – c. CE 300*. Berlin: De Gruyter.

Sijpesteijn, Petra. 2007. 'Palaeography', in Kees Versteegh *et al.* (ed.), *Encyclopedia of Arabic language and linguistics*, vol. 3. Leiden and Boston: Brill, pp. 513–524.

Spitzmüller, Jürgen. 2013. *Graphische Variation als soziale Praxis: eine soziolinguistische Theorie skripturaler 'Sichtbarkeit'*. Berlin: De Gruyter.

Stolk, Joane V. 2015. *Case variation in Greek papyri. Retracing dative case syncretism in the language of the Greek documentary papyri and ostraca from Egypt (300 BCE – 800 CE)*. PhD thesis, University of Oslo.

Swain, Simon. 1996. *Hellenism and empire: Language, classicism, and power in the Greek world, AD 50–250*. Oxford: Clarendon Press.

Tabachovitz, David. 1943. *Études sur le grec de la basse époque*. Uppsala: Almqvist and Wiksell.

Teodorsson, Sven-Tage. 1977. *The phonology of Ptolemaic Koine*. Göteborg: Acta Universitatis Gothoburgensis.

Thompson, Edward M. 1912. *An introduction to Greek and Latin palaeography.* Oxford: Clarendon Press.

Turner, Eric G. 1977. *The typology of the early codex*. Philadelphia: University of Pennsylvania Press.

Turner, Eric G. 1978. *The terms recto and verso. The anatomy of the papyrus roll*. Brussels: Fondation égyptologique Reine Élisabeth.

Vandorpe, Katelijn. 2009. 'Archives and dossiers', in Roger S. Bagnall (ed.), *Oxford handbook of papyrology*. Oxford: Oxford University Press, pp. 216–255.

Vierros, Marja. 2012. *Bilingual Notaries in Hellenistic Egypt*. Brussel: Koninklijke Vlaamse Academie van België voor Wetenschappen en Kunsten.

Walker, Sue. 2001. *Typography and language in everyday life*. Harlow and New York: Longman.

Winter, Werner. 1999. 'Sociolinguistics and dead languages', in Ernst H. Jahr (ed.), *Language change. Advances in historical sociolinguistics*. Berlin and New York: De Gruyter, pp. 67–84.

Wipszycka, Ewa. 2009. *Moines et communautés monastiques en Égypte (IVe–VIII siècles)*. Varsovie: Warsaw University, Raphael Taubenschlag Foundation.

Youtie, Herbert C. 1973. *Scriptiunculae*. Amsterdam: A.M. Hakkert.

Yuen-Collingridge, Rachel and Malcolm Choat. 2012. 'The Copyist at work: scribal practice in duplicate documents', in Paul Schubert (ed.), *Actes du 26e Congrès international de papyrologie. Genève 16–21 août 2010*. Geneva: Droz, pp. 827–34.

7. Spoken Greek and the Work of Notaries in the Acts of the Council of Chalcedon

Tommaso Mari

1 Introduction

The church councils of late antiquity were summoned to discuss and deliberate on important matters of the Christian doctrine and church governance. Hundreds of clerics and occasionally dozens of lay officials attended such councils. The oral medium was predominant and the recording of the proceedings was crucial; for each council, acts were produced that were consulted in the years to come. We have acts of several councils from late antiquity, and the minutes of some of these are very extensive. Since these present themselves as the verbatim transcripts of the conciliar proceedings, they offer an unrivalled insight into the history of major events of late antiquity and into the spoken language of this period (cf. Millar 2006: 16, 249–50). What we can make of these documents for our purposes depends very much on their reliability, which depends in turn on how faithfully they were recorded and transmitted – hence the importance of the work of the notaries (*notarioi*), who were tasked with the production of the records.[1]

[1]Following a now established tradition of scholarship on the Acts, throughout this article I shall use 'notary' as a translation of Latin *notarius* and Greek *notarios*. However, I should make it clear that while English 'notary' is convenient and etymologically consistent, it is inaccurate with regard to the function of *notarii* at this time: they acted as minute takers and secretaries, not as modern notaries (see Teitler 1985 for an overview of these figures). In fact, 'notary' is not contemplated as a translation of *notarius* in the *OLD*, which gives instead 'short-hand writer, stenographer'; the Revised Supplement to the *LSJ* s.v. *notarios* has 'Lat. *notarius*, secretary', and Lampe has 'shorthand writer, secretary' (I should like to thank Hans Teitler for discussing this with me in a private exchange).

In this Section, I shall focus on the minutes included in the Acts of the Council of Chalcedon (451 CE).[2] This council marked a turning point in the history of the church and in late antique history more generally. Hundreds of bishops were summoned, mostly from the Greek East, to produce a new definition of faith and to assess the events of the controversial Second Council of Ephesus (August 449). The latter task required the minutes of Second Ephesus to be read out and discussed. At Second Ephesus, chunks of the minutes of previous gatherings had been read out and discussed, too: the First Council of Ephesus (431), the Resident Synod (*synodos endemousa*) of Constantinople held in November 448 and some related hearings held at Constantinople in April 449 (see Price and Gaddis 2005: I.113–14). Parts of these made it into the Acts of the Council of Chalcedon. This process tells us a good deal about the importance of minutes in this sociocultural context.[3]

In what follows, I shall look at the work of the notaries of the Council of Chalcedon and of the above-mentioned gatherings, in order to discuss the historical and especially linguistic reliability of the minutes;[4] I shall subsequently explore the potential of the minutes as evidence for the spoken Greek of the mid-fifth century by investigating the differences between spoken and written language in the Acts.[5]

I should like to make it clear that I shall concern myself with spoken language as a linguistic medium and not with orality as a linguistic conception, although

---

[2] The Acts are published in Schwartz (1933–7). The Greek Acts are contained in volume II.1, the Latin translation in II.3. An English translation is in Price and Gaddis (2005). When quoting from the Acts (*ACO* = *Acta Conciliorum Oecumenicorum*), I shall indicate volume, page and line number in Schwartz's edition, alongside the number of the session and paragraph (e.g. *ACO* II.1 p. 55.1–6, 1.1). I should point out that Price and Gaddis (2005) in numbering the sessions follow the Latin version, which is at times different from the Greek one (see Price and Gaddis 2005: II.vii–viii).

[3] See Graumann (2009) on aspects of the reading of documents and sets of conciliar acts at First Ephesus but also at Chalcedon. Graumann (2018) focuses on the material objects containing conciliar acts and their archival preservation.

[4] I shall not concern myself with the Acts of First Ephesus, which are preserved independently of the Acts of the Council of Chalcedon (unlike those of the other councils here considered) and pose a different set of problems (see e.g. Graumann 2009).

[5] Our source will be obviously the extant Greek Acts. The original Acts included the Latin statements of the few western delegates, accompanied by a Greek translation (see Schwartz 1933: 247–8); at least some Latin statements must have been still accessible by the time the Latin translations were produced in the mid-sixth century (see Mari 2018). In the extant Greek Acts, all text in Latin has been eliminated.

there obviously is an interplay between these two.[6] As councils were formal occasions attended by bishops and high-ranking imperial officials, the spoken language we must expect to find is by and large that of educated men expressing themselves at a formal occasion, that is formal spoken language.[7]

2 From the oral discussion at the councils to the modern edition of the Acts: a hypothesis

First of all, we need to address the question as to how the minutes were produced. Unfortunately, the Acts of the Council of Chalcedon do not contain information about the minute-taking at that council. Normally, details of this were meant to be invisible, and most of the times they actually remained so unless issues were raised about the veracity of the minutes at following gatherings (Price and Gaddis 2005: I.75–6). For example, the veracity of the minutes of the Resident Synod of 448 came into question at the hearings of Constantinople in 449, and that of the minutes of Second Ephesus in 449 came into question at Chalcedon. Through this scrutiny, some details of the minute-taking in both occasions were revealed, and we shall look at these in what follows.

The most detailed, albeit not entirely clear, source about minute-taking and production of acts at a church gathering are the Acts of the so-called Conference (*Collatio*) of Carthage in 411.[8] Here two groups of rival bishops, the Catholics and the Donatists, had four notaries each (*notarii ecclesiastici*). Two of them for each side would alternate in taking shorthand notes of the proceedings, assisted by a team of imperial stenographers (*exceptores*) and supervised by two imperial *scribae*. The formal version of the minutes was produced after each shift by comparing the parallel versions of the shorthand notes, under the supervision of some representatives of the bishops who had to verify and sign it. After this, the imperial *exceptores* would retranscribe the verified minutes producing the final official version, from which authenticated copies would be made for the different parties.

---

[6] On this distinction and on questions of orality in text (especially Latin texts), see Oesterreicher (1997); although he does not take conciliar acts into account, these would fit most naturally in the text type called 'records of spoken transactions' (Oesterreicher 1997: 202–3). For questions of orality in Attic prose, see Vatri (2017: 1–22).

[7] On formal spoken language and its similarities with formal written language, see Akinnaso (1985).

[8] See Lancel (1972: 337–63) and Teitler (1985: 5–15). The Acts of the Conference of Carthage have been republished most recently by Weidmann (2018).

Different assemblies must have had different systems for the production of minutes, depending on their size, chair, location and so on. At the Resident Synod of Constantinople in 448, for example, five notaries of the archbishop Flavian of Constantinople (Aetius, Asclepiades, Asterius, Nonnus and Procopius) were in charge of the minutes;[9] one year later they still possessed the original minutes authenticated by the signatures of the bishops, and they were required to produce these as their reliability came into question at some hearings held in Constantinople in April 449.[10] Copies were in the possession of some representatives of Eutyches, the monk who was condemned for heresy at the Resident Synod in 448 and who questioned the reliability of the minutes after he allegedly found in them some things contrary to the truth;[11] it is not stated who made such copies – indeed in 449 Flavian's chief notary Aetius asked to inspect them to figure out whose hand they were in and who had provided them, but his request was not granted.[12] Aetius also asked whether the minutes produced by Eutyches' representatives (Constantine, Eleusinius and Constantius) were originals, copies or else; this might suggest that there could have been more than one original, since he too possessed the original minutes.[13] Three *exceptores* are recorded as attending the

---

[9] See Teitler (1985: 108 s.v. Aetius 2, 114 s.v. Asclepiades, 115 s.v. Asterius 4, 154 s.v. Nonnus, 163 s.v. Procopius 3). All of these are called 'deacons and notaries' (*diakonoi kai notarioi*) in the minutes. They also acted as secretaries, reading out documents, answering questions about the proceedings and making announcements.

[10] *ACO* II.1 p. 154.7–8 (session 1 para. 588), an official tells the notaries: Τῶν ὑπομνημάτων τῶν αὐθεντικῶν χρεία ἐστίν, ἐν οἷς αἱ ὑπογραφαὶ τῶν ἐπισκόπων περιέχονται ('We need the original minutes in which the signatures of the bishops are contained'). At *ACO* II.1 p. 156.5 (1.614) it is said that the notaries presented τὸ αὐθεντικὸν σχεδάριον ('the original draft'), which Teitler (1985: 102) says is the equivalent of the *scheda*, the verified and signed minutes, of the Conference of Carthage in 411 (see also Graumann 2018: 284–9). At Second Ephesus (August 449), the original minutes of the Resident Synod of 448 were presented by Archbishop Flavian of Constantinople (cf. *ACO* II.1 p. 99.29–30, 1.222).

[11] Eutyches' petition to the emperors regarding his case read: 'For yesterday I read the minutes that the most devout bishop Flavian has mischievously prepared against me and I found in the text things that are contrary to the proceedings. For neither what he has said to me was contained in it nor did they put down in the minutes what I said' (*ACO* II.1 p. 152.24–7, 1.572).

[12] *ACO* II.1 p. 155.32–7 (1.610–11).

[13] *ACO* II.1 p. 155.26–7 (1.606): πότερον αὐθεντικά ἐστιν ἢ ἀντίγραφα ἢ τί τοιοῦτο παρά τινος αὐτῶι παρεσχέθη ('whether they are originals or copies or what someone ever gave them'). The question was slightly rephrased by another attendee, the patrician Florentius: Ὁ εὐλαβέστατος Κωνσταντῖνος εἰ ἃ προφέρει ἀντίγραφα, ἴσα εἰσὶν ἢ αὐθεντικά, διδάξει ('the most devout Constantine will show if the copies that he is presenting are replicas or originals'); Constantine

hearings in 449 and reading out documents: Asterius, Euethius and John.[14] There is evidence that at Constantinople in 449 the first notes were taken on tablets, for a statement of the deacon Eleusinius was read out *apo deltōn* 'from tablets' shortly after he made it.[15] Hence Teitler (1985: 103) argues, by analogy with the Conference of Carthage in 411, that shorthand symbols were used by the notaries of Flavian of Constantinople as well as by the three *exceptores* who attended the hearings in 449, although explicit evidence is lacking.

We know less about the minute-taking at the Second Council of Ephesus in 449, for which our information comes from complaints raised two years later at the Council of Chalcedon.[16] Dioscorus, bishop of Alexandria, presided over this council, and the protonotary (*protos notariōn*) John of Alexandria was in charge of reading out documents. Several bishops had their own private notaries but claims were made at Chalcedon that Dioscorus expelled the other bishops' notaries and had his own take care of the minutes. Dioscorus tried to defend himself by pushing the idea that each notary took records for his bishop.[17] Bishops Juvenal of Jerusalem and Thalassius of Caesarea, Dioscorus' allies at Second Ephesus, confirmed that they too had their own notaries;[18] however, Dioscorus later revealed that it was his notaries in particular who had taken the minutes, by letting it slip that his notary Demetrianus had been secretly asked by Basil, bishop

---

replied that they were replicas (Ἴσα ἐστίν.) (*ACO* II.1 p. 155.28–30, 1.607–8). At Second Ephesus in 449, Eutyches produced presumably the same replicas: ἐπιδέδωκεν δὲ τὰ ἴσα καὶ ὁ θεοσεβέστατος ἀρχιμανδρίτης Εὐθυχής (*ACO* II.1 p. 99.30–31, 1.222).

[14] See Teitler 1985: 115 s.v. Asterius 5, 132 s.v. Euethius 2, 144 s.v. Iohannes 5, respectively.

[15] *ACO* II.1 p. 169.7 (1.741).

[16] *ACO* II.1 p. 87.10–88.4 (1.122–30).

[17] *ACO* II.1 p. 87.16–20 (1.124): Ἕκαστος διὰ τῶν ἑαυτοῦ νοταρίων ἔγραψεν, οἱ ἐμοὶ τὰ ἐμά, οἱ τοῦ θεοσεβεστάτου ἐπισκόπου Ἰουβεναλίου τὰ αὐτοῦ, οἱ τοῦ θεοσεβεστάτου ἐπισκόπου Θαλασσίου τὰ αὐτοῦ· ἦσαν δὲ καὶ ἄλλων εὐλαβεστάτων ἐπισκόπων πολλοὶ νοτάριοι ἐκλαμβάνοντες. οὕτως οὐκ ἔστιν τῶν ἐμῶν νοταρίων τὸ γράμμα· ἕκαστος ἔχει τὸ ἴδιον ('Each one wrote through his own notaries: mine wrote my records, those of the most religious Juvenal wrote his, those of the most religious bishop Thalassius wrote his; there were also many notaries of other most devout bishops who kept a record. So the text is not of my notaries; each has his own.'). Price (in Price and Gaddis 2005: I.152–3) translates οἱ ἐμοὶ τὰ ἐμά etc. as 'mine recorded my statements' etc., but I find it hard to believe that each notary was only recording the words of his own bishop, for it would have been ultimately pointless if these were not inserted in the context of the debate. I think it more likely that τὰ ἐμά implies ὑπομνήματα ('records') or γράμματα ('texts').

[18] *ACO* II.1 p. 87.21–7 (1.124–7).

of Seleucia, to modify his statement.[19] Some of Dioscorus' accusers repeatedly claimed that they had to sign blank papers.[20]

The Council of Chalcedon was directed by imperial authorities. All sessions but one were chaired by imperial officials, and two imperial secretaries (*sekretarioi*), Constantine and Veronicianus, were tasked with reading out written texts. The patriarchal staff of Constantinople cooperated: the aforementioned Aetius, who had been promoted to Archdeacon of Constantinople and chief of the notaries (*primikerios notariōn*), helped read out documents at some sessions and must have played a role in checking the minutes.[21] The third session, Dioscorus' trial, was exceptional:[22] it was presided over by the chief of the Roman delegation, Paschasinus bishop of Lilybaeum, and documents were read out by the patriarchal notaries Aetius, Asclepiades and Procopius (Asclepiades acted as reader alongside Aetius at the fifth session as well); moreover, three delegations of bishops were sent to Dioscorus with summons, and each of them included one lector and notary (*anagnostēs kai notarios*) who took notes and read them back before the assembly.[23] At Chalcedon like at Second Ephesus some bishops came with their own notaries.[24] Some version of the minutes was ready for use soon after the sessions.[25]

The Conference of Carthage in 411 was special in many ways: although it was presided over by a delegate of the emperor, like the Council of Chalcedon,

---

[19] *ACO* II.1 p. 179.37–180.2 (1.854). For Demetrianus, see Teitler (1985: 127 s.v. Demetrianus).

[20] *ACO* II.1 p. 88.5–16 (1.131–4).

[21] In the Latin Acts, the first session is concluded by Aetius' statement 'It is complete' (*ACO* II.3 p. 259.18–19, 1.1076); this must be 'a record of a subsequent checking of the minutes' (Price and Gaddis 2005: I.365 n. 523).

[22] This session was the third chronologically and in the Latin Acts, but it is numbered as the second in the Greek Acts, for the order of the sessions was rearranged.

[23] Himerius and Hypatius, lectors and notaries, and Palladius, deacon and notary of Patricius bishop of Tyana (see Teitler 1985: 141 s.v. Himerius, 142 s.v. Hypatius 2, 156 s.v. Palladius 5).

[24] *ACO* II.1 p. 78.8–11 (1.75–6): the two notaries of Dioscorus of Alexandria are accused of being rowdy.

[25] For example, at the fourth session (17 October 451) the secretary Constantine read out parts of the minutes of the first session (8 October) and the secretary Veronicianus parts of the minutes of the second session (10 October) (*ACO* II.1 pp. 288–9, 4.2–4); at the seventeenth session (sixteenth in the Latin Acts and in Price and Gaddis 2005), the minutes of a private meeting that had been held the previous day were read out by Aetius (*ACO* II.1 pp. 447–53, 17.7–9). In all three of these cases the term used to indicate the object containing the minutes is *schedarion* (see Graumann 2018: 284–5).

it was more of a show trial than a council and was 'a formally and explicitly adversarial affair between two separate churches whose bishops deeply mistrusted each other' (Price and Gaddis 2005: I.75). It must have been a consequence of this that two teams of notaries took shorthand notes in parallel and that representatives of the bishops supervised the production of the formal minutes and verified them with signatures. It is unlikely that this particular procedure was put in place at other gatherings; yet it stands to reason that the more basic aspects of the production of the minutes (i.e. notaries taking shorthand notes and later rendering them into formal minutes) were the same at the Conference of Carthage and at the Council of Chalcedon – as well as at those gatherings whose minutes made it into the Acts of the Council of Chalcedon. We may thus reconstruct the stages from the oral discussion at the councils to the modern edition of the Greek Acts as in Table 1:

Table 1.

Stage 1	Spoken statements and written texts (read out)
Stage 2	Notaries produce shorthand transcription during the sessions
Stage 3	Notaries render shorthand transcription into formal version
Stage 4	Copies of the formal version are made for parties[26]
Stage 5	Official publication of the Acts of Chalcedon (Constantinople, 454/455)[27]
Stage 6	Revision of the Greek Acts (probably seventh century)[28]

---

[26] We know that the Roman delegates returned to Rome from Chalcedon with some documents of the Council, and Anatolius of Constantinople sent the rest of the minutes to Pope Leo by December 451 (Letter of Anatolius to Leo, ACO II.1 p. 448.24–8).

[27] See Price and Gaddis (2005: I.79–81). The publication of the Acts was promoted by the imperial court and patriarchal see of Constantinople as a means of propaganda; the Acts include not only the minutes of the proceedings but also letters and other documents related to the Council of Chalcedon.

[28] Price and Gaddis (2005: I.82–3). In the sixth century, three Latin translations of the Acts had been produced (see Price and Gaddis 2005: I.83–5 and Mari 2018); the extant Greek version is sometimes less complete than the Latin translations, for it suffered cuts after the Latin translations were made.

Stage 7	Medieval manuscript tradition[29]
Stage 8	Schwartz's critical edition (1933–7)

A clarification on the first stage is in order. Here I take spoken statements to include both unprepared statements and oral speeches that were based on written texts, for they function in the same way from the perspective of the production of the minutes: both types of speech were recorded as they were delivered.

By written texts I mean petitions, letters and minutes of previous gatherings that were read out at the councils. From the perspective of the production of the minutes, these might differ from spoken statements, for it is conceivable that they could be handed to the notaries to be copied instead of being transcribed as they were read out.[30]

Spoken statements and written texts will constitute the basis of my comparison between spoken and written language in Section 4. I should make it clear that, for the purposes of the linguistic analysis, spoken statements included in the minutes of previous gatherings count as evidence of spoken language, not of written language. I should add that spoken statements based on written texts, if they are not recognised as such, may cause problems in the analysis of spoken language (see Section 4.2).

Now that we have an idea of how the minutes have been produced and transmitted, we can turn to the question of their historical and especially linguistic reliability. First, how faithfully do the minutes report the contents of the gatherings? Second, and most crucially for us, how faithfully do they represent the language spoken at the gatherings? The second question depends to some extent on the first, inasmuch as alterations in the contents of minutes would produce alterations in their language, thus undermining their faithfulness to the language spoken at the gatherings.

---

[29] Schwartz's edition is mainly based on the manuscripts of Venice, Biblioteca Nazionale Marciana, Gr. Z. 555 (eleventh century) and Vienna, Österreichische Nationalbibliothek, hist. gr. 27 (twelfth/ thirteenth century).

[30] This seems to be the meaning of requests that frequently accompany the reading out of written texts, such as 'let this be read and inserted in the text of the minutes' (e.g. *ACO* II.1 p. 83.22–3, 1.86; p. 90.13–15, 1.156; p. 100.12–13, 1.223, etc.).

3 The Acts as historical evidence

In this Section, I shall discuss the Acts as historical evidence; I shall use as a basis for discussion the work of Price (2009), who has investigated the question of how much in the Acts of the Council of Chalcedon belongs to the categories of truth, omission and fiction, respectively. He has convincingly argued that 'the first, fortunately, greatly outweighs the third', and also that 'the category of omission is much more significant than that of fiction' (Price 2009: 105). On this plausible conclusion I shall elaborate in what follows. It is beyond the scope of this contribution to assess the historical reliability of the Acts altogether; my focus will be on the recording policies and practices of the notaries (with the proviso that they might have been different at different gatherings), and on the significance that these might have for our linguistic appreciation of the Acts.[31] I shall discuss the categories identified by Price (2009) in the following order: omission, fiction (which I call 'alteration and falsification') and, by process of elimination, truth.

3.1 Omission

Generally speaking, omissions are more likely to undermine the evidentiary value of a document for historical than for linguistic investigation. From a linguist's point of view, omissions simply reduce the size of the corpus, unless they target linguistically marked material, in which case they weaken the representativity of the corpus (e.g. if the records of a meeting were to omit all statements of those speaking a substandard variety of the language).

The most striking example of omission at Chalcedon is that of sessions that were not recorded at all (cf. Price 2009: 97–8).[32] For example, an unrecorded meeting in the palace of the archbishop Anatolius of Constantinople was meant to convince everybody of the orthodoxy of Pope Leo's main theological work, the so-called *Tome*;[33] in this way, there would have only been consensus at the

---

[31] On contemporary challenges to the accuracy of the records see Ste. Croix (2006: 307–10).

[32] There were also extra-conciliar meetings, such as that in 448 between some envoys of the Resident Synod and the monk Eutyches, of which only informal notes were taken and whose content was then reported before the assembly. Upon being questioned about his account of the meeting, the presbyter John admitted that 'it is not possible for one who is sent to convey a message to others to report back the exact words' (*ACO* II.1 p. 160.5–7, 1.644).

[33] *ACO* II.1 p. 279.8–11 (3.33; 2.33 in the Latin Acts and in Price and Gaddis 2005).

formal session in which Leo's Tome was to be approved.[34] If minutes had been taken at that 'private' meeting, we would certainly know more about the degree of doctrinal dissent over Leo's Christology, but the records would hardly give us a very different picture of the language than we can find at other sessions. The same must be true of material that was excised later for the sake of brevity,[35] and material that was omitted for the sake of convenience.

The very different length of some sessions is striking. For example, the first session, where the events of Second Ephesus were assessed, spans 142 pages, as opposed to only ten pages for the crucial fifth session, at which the draft definition of faith was read out, discussed and amended. We know that the first session was exceptionally long and went on until late at night, but it is clear that much of the fifth session has been omitted. In some cases that is explicit in the very minutes: at paragraph 3, we read that 'Asclepiades, deacon of the great church of Constantinople, read out the definition, which it was decided not to include in these minutes' (*ACO* II.1 p. 319.7–8); at paragraph 4, it is stated that some people raised objections after the reading, but not who did it and what objections were raised (*ACO* II.1 p. 319.9); at paragraph 29, it is reported that a selected committee met to discuss the amendments to the definition, but the discussion itself is not recorded (*ACO* II.1 p. 322.1–2). Moreover, the objections to the draft definition of John bishop of Germanicia (paragraph 4) and of the Roman delegates (paragraph 9) must have been much more detailed than we read now to justify the long and animated responses attributed to 'the most devout bishops' at paragraphs 6, 11, 12, etc.[36]

Who decided what was not to be included in the minutes, and based on what criteria? In the case of the draft definition (*ACO* II.1 p. 319.7–8, 5.3), the phrasing suggests that the decision to omit the draft definition was taken during the production of the formal minutes; the reason for this omission as well as the omission of the objections to the draft definition was probably that the editors did not want to provide arguments to the critics of the definition of faith (Price and

---

[34] Deliberative processes at councils depended on a system of unanimity, not majority (see Ste. Croix 2006: 266–7 and Price 2009: 92–5).

[35] For example, the reading of the Acts of First Ephesus spans 40 pages in the Latin version (*ACO* II.3 pp. 196–235) but it is reduced to a very short summary in the Greek version (*ACO* II.1 p. 189.31–4). Just after that, where the Acts of Second Ephesus are read out, the Greek version omits most of the sentences of the bishops and only gives their names (*ACO* II.1 p. 190.15–22, 1.945–51, and p. 190.29–33, 1.954–7), while the Latin version preserves the full sentences (*ACO* II.3 pp. 236.4–237.4, 1.948–54, and p. 237.10–24, 1.957–60). These parts must have been omitted in the seventh-century revision of the Greek Acts for the sake of brevity.

[36] See Price (2009: 96–7) for more examples; see also Ste. Croix (2006: 266, 300).

Gaddis 2005: I.196 n. 33). The aim to portray ecclesiastical consensus instead of disagreement certainly played a role: this must have been especially the case with sessions focusing on doctrine, such as the fifth, while there was an interest to record the 'judicial' sessions more fully (Price and Gaddis 2005: I.78).[37]

But what is more relevant for our linguistic investigation is that it also seems that a criterion of formality played a role in the selective recording of the proceedings: to this effect we have the testimony of Aetius, the chief notary at the Resident Synod of Constantinople in 448. As I have said in Section 2, at this Resident Synod, the archimandrite Eutyches was condemned for heresy; he later appealed, claiming that the minutes of the synod had been falsified, and some hearings took place in April 449 to reexamine the minutes. Eutyches did not attend this meeting in person but was represented by three monks: Constantine, Eleusinius and Constantius. They possessed copies of the minutes and checked them against the official minutes as they were being read out by an *exceptor*. At one point, Constantine observed that three statements of Flavian of Constantinople,[38] Seleucus of Amaseia and Basil of Seleucia were missing from the official minutes (*ACO* II.1 p. 172.34–173.10, 1.788); the notary Aetius quite candidly replied that 'many things are often said in the way of ordinary conversation and suggestions (ὡς ἐν διαλέξει κοινῆι καὶ ἐν συμβουλῆς μέρει) in synod by the most holy bishops present that they do not command to write down (ἃ οὐκ ἐπιτρέπουσι

---

[37] Not that omissions cannot be identified in judicial sessions such as the first, though: for example, the oriental bishops' exclamation as Bishop Theodoret was admitted to the council, 'we signed blank sheets. we were beaten and we signed' (*ACO* II.1 p. 69.21, 1.28), can only be explained if somebody just accused them of having previously signed his condemnation, which is not in the minutes (cf. Price and Gaddis 2005: I.134 n. 66). As for the hearings at Constantinople in 449, Eutyches' defensive strategy was based on the claim that some statements had been omitted in the minutes of the Resident Synod of Constantinople in 448, where he had been deposed ('For neither what he [Bishop Flavian] has said to me was contained in it nor did they put down in the minutes what I said' *ACO* II.1 p. 152.24–7, 1.572): so his delegates at the hearings in April 449 lamented omissions at *ACO* II.1 p. 168.30–4 (1.737), p. 171.28-31 (1.773), p. 172.34–173.10 (1.788), p. 174.8 (1.797), p. 174.26–8 (1.804), p. 175.30–32 (1.818). Some of these claims were refuted by other attendees, so it is difficult to tell what was actually omitted and what was never said at all. A good deal of omission must have affected the minutes of Second Ephesus in 449, that were controlled by the notaries of Dioscorus; some bishops at Chalcedon recalled statements and events that are not recorded in the minutes of Second Ephesus and against which Dioscorus protested strongly: *ACO* II.1 p. 180.3–9 (1.855), p. 180.14–28 (1.858), p. 180.33–40 (1.861), etc.

[38] The statement attributed to Flavian of Constantinople ('say "two natures after the union" and anathematise those who do not say so') must have been made before Eutyches' statement at *ACO* II.1 p. 143.32 (1.535); as it was not commented upon immediately, Constantine brought it up again at *ACO* II.1 p. 174.25–175.29 (1.804–17).

γράφεσθαι)' (*ACO* II.1 p. 173.32–4, 1.792). There is an implicit distinction here between formal pronouncements and informal communication: the former were to be recorded, the latter not. After Basil replied that he did not remember exactly what he had said and reconstructed something different from what Constantine found in the minutes (*ACO* II.1 p. 173.19–31, 1.791), Constantine repeated that that was not contained in the minutes. Basil then admitted having said it but in conversation and not as a declaration (διαλεγόμενος τότε, οὐκ ἀποφαινόμενος), implying that that was the reason why his statement was not recorded (*ACO* II.1 p. 174.9–13, 1.798).[39] This situation recurs several times at the hearings of Constantinople in 449: after the *exceptor* read out a statement of the patrician Florentius ending with 'Speak!', Florentius complained that he said 'Speak!' not as a pronouncement (ὡς διαλαλῶν) but as an exhortation (προτρέπων), evidently implying that that word should not have been put down in the records (*ACO* II.1 p. 171.25–7, 1.772);[40] Florentius brought up the same complaint again about a slightly longer exhortation of his ('I did not say "Speak! If you do not speak, you are deposed" as a pronouncement', *ACO* II.1 p. 172.1–3, 1.776).

The statement of Seleucus of Amaseia that was found in Constantine's version of the minutes but not in the official version was by and large confirmed by the patrician Florentius and by Seleucus himself (*ACO* II.1 p. 173.11–18, 1.789–90); the notary Aetius asked Seleucus if his statement was meant to be included in the minutes, and Florentius followed up on that by asking if Seleucus or anybody else said what needed to be recorded and what did not (*ACO* II.1 p. 173.11–18, 1.799–800); Seleucus remembered having said such things (μέμνημαι εἰρηκὼς τοιαύτας φωνάς) but blamed the failure to record his statement in full on the uproar that followed (*ACO* II.1 p. 174.23–4, 1.803).

So apparently it was not only up to the notaries to work out what was meant as a formal pronouncement and what was ordinary conversation; the attendees could have their say in asking that some utterances be recorded or not. We often come across speakers explicitly asking that some testimonies or written documents be included in the minutes;[41] what is more striking, some

---

[39] This statement of Basil was discussed also at Second Ephesus (*ACO* II.1 p. 144.28–145.4, 1.546–8) and at Chalcedon (*ACO* II.1 p. 92.18–93.2, 1.168–9).

[40] See Price and Gaddis (2005: I.258 n. 295), following Schwartz (1929: 30).

[41] This especially concerns written documents (all from *ACO* II.1): p. 83.22–3 (1.86) and p. 90.13–15 (1.156) at Second Ephesus; p. 102.22–9 (1.235), p. 104.5–7 (1.238), p. 126.12–16 (1.378–9) at the Resident Synod of 448; p. 147.34–5 (1.554) at Second Ephesus regarding the minutes of the hearings of Constantinople in 449. Sometimes it is also asked that oral testimonies or statements be included in the minutes: p. 137.13–15 (1.457) at the Resident Synod of 448; p. 176.34–6 (1.828),

even asked that their statements be deleted from the minutes. This is the case of the monk Constantine at *ACO* II.1 p. 156.28–157.22 (1.621–8): he first asked that an ill-judged comment of his (1.621) be erased, because he allegedly made it during an uproar without being aware (1.624); his request was not granted: the bishop Seleucus replied that the comment was made in a quiet moment before the uproar (1.625), while Thalassius and Eusebius stated that Constantine could not be selective about his own statements but had to accept all of them (1.627–8). Constantine did not go quietly, and kept on insisting that he said one word during an uproar and that was recorded (*ACO* II.1 p. 158.20–1, 1.639). As a matter of fact, uproars were anything but exceptional at councils;[42] there is evidence that they made it difficult for speakers to express themselves clearly, for listeners to hear and, we must assume, for notaries to take records accurately. For example, when the bishops at the hearings of Constantinople in 449 were asked whether they heard Flavian's statement as found in Constantine's minutes but not in the official minutes (1.805), Basil, Julian and Longinus replied that they could not remember due to the uproar (*ACO* II.1 p. 174.29–175.26, 1.808, 814, 816); when Constantine pointed out that, after his deposition, Eutyches made an appeal that was not recorded in the official minutes (1.818), the patrician Florentius replied that Eutyches said that to him softly (*praōs*) during an uproar after the closing of the synod (1.819), and other bishops stated that they never heard Eutyches say that (*ACO* II.1 p. 175.30–176.10, 1.818–24).

To sum up, we have seen that informal statements were not meant to make it into the records, and they generally did not; however, some did. It is a difficult question whether omissions of informal statements have affected the linguistic representativity of the Acts as a whole by leaving out linguistically marked material. The councils were formal situations with a quite well-defined procedure, and there is some degree of formulaicity in the language of the attendees in certain occasions (e.g. when they express themselves on doctrine). It may be, although it need not be, that informal statements, which have not been recorded, contained more informal linguistic features than formal pronouncements, which have been recorded.

3.2 Alteration and falsification

For the purposes of our investigation, I take Price's category of fiction to include those statements included in the minutes whose wording has been altered

---

p. 178.29–30 (1.843) and p. 179.7–9 (1.847) at the hearing of Constantinople in 449.

[42] On unruly behaviour at councils, see Whitby (2009).

(deliberately or not) or that have been made up entirely. Surely, while omissions merely reduce the material available for our linguistic analysis, such instances of alteration and falsification would more deeply undermine the value of our corpus as evidence for spoken language; for some statements that we look at as samples of spoken language might have been actually rewritten or written in the first place. Forgery was a hot issue at councils (see Wessel 2001). As a matter of fact, most claims of forgery in the Acts concern omissions, not additions or alterations; this is especially the case with the minutes of the Resident Synod of 448 as examined at the hearings of Constantinople in 449. That is in keeping with Price's (2009: 105) conclusion that 'the category of omission is much more significant than that of fiction'. For example, the patrician Florentius complained that the minutes of the Resident Synod of 448 ascribed to him two sentences that he had never uttered (*ACO* II.1 p. 167.1–6, 1.721; p. 172.11–12, 1.778); in both cases his complaints got the notaries in some trouble, for they prompted Archbishop Flavian of Constantinople to question them insistently about their work (p. 167.6–14, 1.722–5; p. 172.13–23, 1.779–81). Constantine, the monk representing Eutyches, lamented inaccuracies in the minutes a couple of times (*ACO* II.1 p. 156.21–3, 28–30, 1.619, 621), which he took back as soon as he realised that it was counter-productive, and the deacon Eleusinius referred that the minutes did not report in the proper order what had happened (p. 167.19–23 (1.728)). The reading of the first two sessions of the Resident Synod of Constantinople in 448 did not spark protests (*ACO* II.1 p. 156.13–18, 1.616–7), and in another couple of cases the representatives of Eutyches had to admit that the minutes were correct (*ACO* II.1 p. 165.3–4, 1.690; p. 166.32, 1.718). At the end of the hearing on the case of Eutyches (13 April 449), the notary Aetius happily concluded that after many readings of the minutes, nobody had found fault with him and the other notaries (*ACO* II.1 p. 176.27–9, 1.827); this did not prevent the notary Asterius from accusing them of having altered certain chapters of the minutes, as the official Macedonius reported (*ACO* II.1 p. 179.1–6, 1.846).

Claims that some statements had been falsified were more frequent at Chalcedon with regard to Second Ephesus. As we have seen in Section 2, Dioscorus controlled the proceedings at Ephesus and it was later alleged that his notaries were in charge of the minutes, while the notaries of those bishops who were not on his side suffered violence and were prevented from taking notes, and those very bishops were forced to put their signatures on blank sheets. It is difficult to say to what extent these claims were truthful and to what extent they were an attempt of some bishops to justify their support for Dioscorus at Ephesus, at a time when it was no longer convenient to be on his side. For example, in

the minutes of Second Ephesus, the bishop Aethericus denies having uttered the statement in support of Flavian that the minutes of the Resident Synod of 448 ascribe to him (*ACO* II.1 p. 118.20–119.4, 1.308–14); at Chalcedon, on the contrary, he claimed that Dioscorus pressurised him to deny that, which prompted Dioscorus to accuse him of calumny (p. 119.15–30, 1.323–9) (cf. Ste. Croix 2006: 308 n. 110).

As the minutes of Second Ephesus were read back, Dioscorus confidently stated that 'the minutes themselves will reveal the truth' (*ACO* II.1 p. 112.6–7, 1.260), but that did not work very well for him. It was especially collective pronouncements that were contested by those bishops who were opposing him at Chalcedon.[43] Here one is reminded of the notary Aetius confessing a notarial 'secret' at the hearing of Constantinople in 449, namely that 'at these most holy gatherings it often happens that one of the most God-beloved bishops present says something and what is said by one is written down and counted as if it was said by everyone alike. This has happened from the beginnings: for example, when one person speaks, we write "The holy council said".' (*ACO* II.1 p. 170.34–7, 1.767). On that occasion, the patrician Florentius picked up on that with a comment to the effect that individual pronouncements recorded in the minutes could be relied upon, but collective pronouncements could not (*ACO* II.1 p. 171.3–4, 1.768; cf. Price and Gaddis 2005: I.257 n. 294). Collective pronouncements and acclamations are very common in the Acts (see Roueché 2009); Aetius' testimony serves as a warning that some of these might have been pronounced by individuals, not by groups – which has consequences both for our historical and for our linguistic appreciation of the Acts.

3.3 Truth

By process of elimination, we could conclude that anything that was recorded in the Acts (i.e. that was not omitted) and was not falsified falls into the category of 'truth'. Of course we do not have enough independent evidence to confirm the veracity of everything that is on record; also, it is certain that falsification was more frequent than we know from the complaints recorded in the Acts.[44] But it is

---

[43] *ACO* II.1 p. 87.8–9 (1.121); p. 89.22–3 (1.149); p. 140.33–4 (1.496); p. 143.14–19 (1.530).

[44] While in the Acts of the Council of Chalcedon there are recorded complaints about falsification in the minutes of previous gatherings, to my knowledge we lack similar complaints about the proceedings of the Council of Chalcedon itself. Yet one can occasionally find evidence for it, as for example in the suspicious differences between the Greek and Latin versions of the crucial sixteenth session (cf. Price 2009: 100–101).

highly unlikely, or I should say impossible, that most or even much of the content of several hundred pages of Greek text was completely or mostly made up by the notaries.[45] To falsify in their entirety the records of such a sizeable gathering would have been a much more challenging and ultimately less profitable task than to falsify precise sections. Indeed, we have seen in Sections 3.1 and 3.2 that claims of falsification always revolved around single sentences. One may add to this that there is a great deal of realistic elements in the Acts that seem unlikely to have been made up, such as embarrassing dissent and unruly behaviour on the part of the bishops (cf. Whitby 2009 and Price 2009: 94–6); also, interruptions are recorded precisely with explicit captions and sentences left hanging, as in example 1:

> (1) *ACO* II.1 p. 155.19–24, 1.604–5
> Ἀέτιος διάκονος καὶ νοτάριος εἶπεν· Εἰ κελεύει ἡ μεγαλοπρέπεια ὑμῶν, ἔχομέν τι εἰπεῖν. ἐμάθομεν ὡς διὰ τῶν δεήσεων ἀνεδίδαξεν ὁ εὐλαβέστατος Εὐτυχὴς πράξει ὑπομνημάτων ἐντετυχηκέναι κἀκεῖθεν τὰς αἰτίας τῶν αὑτῶι προσουσῶν δικαιολογιῶν εὑρηκέναι· ταύτην τὴν πρᾶξιν αξιοῦμεν
> Οὗ λέγοντος ὁ μεγαλοπρεπέστατος πατρίκιος εἶπεν· …
> Aetius deacon and notary said: 'If your magnificence gives permission, we have something to say. We have heard that through his petition the most devout Eutyches declared that he had read the minutes and found there the grounds for his defence. We ask that this text...'
> While he was speaking, the most magnificent patrician said: …

As a rule, of course, verisimilitude is no guarantee of truth; realistic details may be artfully inserted into a forgery so that it does not look like a forgery. But that hardly seems to be the case here, and a healthy scepticism cannot detract from the evidentiary value of the Acts as a historical document.

---

[45] Famously, this is what Riedinger believed happened at the Lateran Council of 649: in his view, the Acts were composed by Greek monks before the council even took place, and the notaries simply read out the script (including the bishops' statements!) during the sessions (cf. Riedinger 1982: 120). However, Price *et al.* (2014: 64–8) have convincingly showed that, while much of the materials must have been planned in advance, there are some elements of spontaneity in the Acts. At any rate, the Lateran Council was very different from that of Chalcedon, for the latter was much longer and involved a great deal of debate, while the former mostly consisted of long and articulate speeches that were quite obviously read out.

4 The Acts as evidence for spoken Greek

I have shown that the Acts are by and large reliable as far as their content is concerned, although they certainly present some problems of omission and, to a lesser extent, falsification. Now we come to our second question, that of the linguistic reliability of the Acts. How faithfully did the scribes record the speakers' utterances from a linguistic point of view? There are several factors to take into account here, some of which we cannot really control. For example, the notaries' skills and the practicality of their writing supports must have played a role; also, the motivations and attitude of the notaries as well as notarial policies are crucial. We have seen in Section 3 that notarial policies were quite thorough but not absolutely so, for notaries were not normally meant to record informal statements. But in recording formal statements, how did they handle less formal features that frequently occur in the spoken language, such as interjections, pauses, repetitions, syntactic inconsistencies, etc.?[46] And how about nonstandard and/or substandard linguistic features, if any?

4.1 A modern parallel: the records of the House of Commons of the United Kingdom

It may be helpful to look at a crosscultural parallel, the official records of a modern deliberative assembly, to appreciate how this can be done nowadays. We have to take into account that now, unlike in the fifth century, sound recording allows one to check minutes and correct any mistakes that may have been made during the first transcription. Potentially, modern records can be a hundred per cent accurate; this makes it all the more significant when they are not so, for any divergences will be the result of choices and, possibly, policies.

I have chosen an Oral Answers to Questions session of the House of Commons of the United Kingdom (22 February 2016).[47] Here is part of Prime Minister David Cameron's impromptu response to Jeremy Corbyn on Britain's EU referendum (3:57pm). First is my verbatim transcription (2), then the official

---

[46] A study of such features in Plato's *Apology* is in Verano (2018). For an account of elements of spoken language in Latin texts, see Koch (1995). Koch and Oesterreicher (2011) discuss spoken language in French, Italian and Spanish. More on this in Section 4.3.

[47] A reference to the records of parliamentary proceedings in the UK is in Ste. Croix (2006: 310 n. 114), where some of the challenges of that process are discussed.

transcription (3). I indicate in bold the differences between the two versions:[48]

> (2) Verbatim transcription of David Cameron's speech
> **Well, let me thank** the right **honourable** Gentleman for his contribution. **Look, he** and I disagree on many, many things **– about** economic policy, **about** social policy, **about** welfare policy**, indeed we even disagree about** the approach we should take within Europe, as he**'s** just demonstrated in his response **–** but we do **both** agree about one thing, **which is that** Britain should be in there, fighting for a good deal for our country. **Erm erm** I worry a little for **the right honourable Gentleman** ... On what **he** said about the (**uh**) deal, **(erm)** I – **I'm going to** make two points about why **I think actually** he should **really** welcome **the deal**. The first is that **it does actually implement**, **as far as I can see**, almost every pledge on Europe in the Labour manifesto – **and** I**'m** looking at the former **(erm) leader.**

> (3) Official transcription of David Cameron's speech
> **I thank** the right **hon**. Gentleman for his contribution. **He** and I disagree on many, many things – economic policy, social policy, welfare policy **and even** the approach we should take within Europe, as he **has** just demonstrated in his response – but we do agree about one thing**:** Britain should be in there, fighting for a good deal for our country. I worry a little for **him** … On what **the right hon. Gentleman** said about the deal, I **will** make two points about why he should welcome **it**. The first is that, **as far as I can see**, **it implements** almost every pledge on Europe in the Labour manifesto – **I am** looking at the former **Labour leader when I say that.**

As is easy to see, the official transcription is a slightly polished version of Cameron's unprepared speech. It eliminates some typical elements of spoken language such as interjections ('erm erm', 'uh'), phatic expressions ('well', 'look'), contractions ('he's' becomes 'he has', 'I'm' becomes 'I am') and repetitions ('I – I'); it adjusts some inconsistencies and stylistic infelicities that must be due to limited time for elaboration (e.g. 'I'm going to make two points about why I think actually he should really welcome the deal' becomes 'I will make two points about why he should welcome it', where the second 'deal' is replaced with the pronoun 'it' and the emphatic adverbs 'actually' and 'really' are omitted); it makes more

---

[48] The video of the session is available online on Parliament TV; the official records of the debate are available online in the House of Commons Hansard.

explicit some expressions that in speech depend on the context and possibly on extralinguistic elements like gestures to be fully understood (e.g. 'and I'm looking at the former leader' becomes 'I am looking at the former Labour leader when I say that'); it eliminates emphatic and/or pleonastic expressions (e.g. 'it does actually implement' becomes 'it implements'; 'indeed we even disagree about the approach' becomes 'and even the approach'; 'agree about one thing, which is that Britain' becomes 'agree about one thing: Britain'), and so on.

While we cannot assume the policies of the stenographers of the House of Commons to have universal value, it stands to reason that some aims and attitudes of fifth-century stenographers were similar. First of all, as is quite obvious, the stenographers of church councils, just like those of the House of Commons, were not interested in the language of the debate but in its content and in the legibility of the minutes. So, we cannot expect the degree of faithfulness that we get in modern transcripts of spoken language recorded for linguistic purposes. For example, we do not find evidence on phonology in the minutes: as the Acts are an official text, the notaries applied the same orthographic conventions that they would have applied in any other official text. Likewise, we cannot expect that they would have recorded interjections such as Cameron's 'erm' or 'uh' and very obvious repetitions like Cameron's 'I – I'. Indeed these do not appear in the sample of the Acts that I have examined (see Section 4.2 for the corpus). There is, however, limited evidence for phatic expressions that are typical of real-time communication; that is the case of Eutyches' parenthetic 'did you pay attention?' at the Resident Synod of Constantinople in 448 (example 4):

> (4) *ACO* II.1 p. 142.26–8, 1.522
> ἐπειδὴ γὰρ σῶμα θεοῦ αὐτὸ ὁμολογῶ (προσέσχες;), οὐκ εἶπον σῶμα ἀνθρώπου τὸ τοῦ θεοῦ σῶμα …
> for since I acknowledge it to be the body of God (did you pay attention?), I did not say that the body of God is man's body …

As in the case of David Cameron's speech, what we can legitimately expect not to have been dramatically altered is the overall syntactic structure of the sentences – unless it was so broken that a reader could not make sense of it – and the lexicon;[49] for to systematically change that would have been a challenging and

---

[49] That the Acts can be a source for colloquial lexicon has been pointed out by Ste. Croix (1984: 23–4), who gives the example of *salgamarioi* 'pickle-sellers', a loanword from Latin (*salgamarius*) which was used in a derogatory sense by Diogenes bishop of Cyzicus (*ACO* II.1 p. 411.30–31, 12.56); this term first made it into the Liddell and Scott in the Supplement to the 9th edition in

time-consuming task, and one that would not have really benefited anyone. Here I wish to suggest that, in order to explore the potential of this material as evidence for spoken Greek, we should look first at syntactic structure and complexity on the one hand and at lexicon on the other hand.

4.2 Spoken and written language: our corpus

The features of spoken language are best appreciated by comparison with written language. In order to analyze the differences between spoken and written language in the Acts, I have put together a corpus of spoken statements and written texts produced at the councils. Spoken statements, which are the vast majority in the Acts, consist of utterances that are not presented as having been read out but that are normally introduced by the verb 'to say'; written texts typically include letters, petitions and bills of indictment and are introduced by the verb 'to read out'.

The samples belong to five of the few attendees who both spoke and presented written texts at the councils. Having a spoken and written set for each of them allows us to investigate how one attendee's spoken language differed from that attendee's written language, thus making up for the impact that idiolects might have on the analysis of the corpus as a whole; it also allows us to compare the language of one attendee with the language of all the other attendees. After comparing the language of individual attendees, I shall attempt to produce generalizations based on the whole corpus without differentiating for individual attendees (Section 4.3).

For each attendee, I have selected a sample of spoken statements and one of written texts that are approximately the same in size. Both the spoken and written samples of Anatolius, Eusebius and Eutyches are a little over 600 words, while the samples of the bishops Bassianus and Photius are smaller, for they spoke and wrote much less than the other three. While I could have chosen to set a cut-off size based on the size of the smallest sample, I have preferred to have larger samples whenever possible so as to increase representativity.

I have aimed for consistency between spoken and written statements with regard to communicative situations, for different communicative situations might require different styles and linguistic features; since the written texts are long and elaborate, I have looked for spoken statements that are also fairly

---

1968, and the only reference given was a Corinthian inscription from the sixth century.

long.[50] Not all attendees delivered long speeches, though: the corpus of Eutyches' oral statements, for example, mostly consists of short answers given during a questioning session. We shall see that different findings correlate with different types of speech. While this detracts from the homogeneity of the corpus as a whole, it also contributes to its diversity and makes it more representative of varieties of actual speech, allowing us to look into different registers. The corpus is represented in Table 2:

Table 2.

Attendee	*Spoken statements*	*Written texts*[51]
Anatolius of Constantinople	**632 words**: *ACO* II.1 p. 206.10–13 (2.12), 225.21–5 (3.95), 290.4–10 (4.9.1), 397.21–4 (11.145), 398.32–399.7 (11.162), 410.37–411.10 (12.50), 412.24–8 (13.3), 413.9–14 (13.9), 466.5–11 (19.32), 468.1–6 (19.50)	**631 words**: *ACO* II.1 p. 248.5–249.21 (letter to Pope Leo)
Bassianus of Ephesus	**390 words**: *ACO* II.1 p. 405.19–406.15 (12.14)	**383 words**: *ACO* II.1 p. 409.1–33 (11.7, petition)
Eusebius of Dorylaeum	**633 words**: *ACO* II.1 p. 66.13–17 (1.14), 103.5–104.7 (1.238), 134.11–23 (1.443), 135.1–11 (1.445)	**636 words**: *ACO* II.1 p. 66.23–67.17 (1.16, petition), 100.18–101.5 + 101.16–28 (1.225 + 230, bill of indictment)

---

[50] See Akinnaso (1985: 330–31) for criticism of studies comparing the two most distant discourse types, formal written language and informal spoken language.

[51] In all of the written texts, I have deliberately left out the salutation formulas at the beginning.

Eutyches of Constantinople	**618 words**: *ACO* II.1 p. 90.7–12 (1.155), 99.17–22 (1.220), 138.33–4 (1.471), 141.5–7 (1.498), 141.12–13 (1.502), 141.20–24 (1.505), 142.4–6 (1.512), 142.8–10 (1.514), 142.13–15 (1.516), 142.18 (1.518), 142.22 (1.520), 142.26–33 (1.522), 143.1–3 (1.524), 143.10–11 (1.527), 143.32–144.2 (1.535), 144.14–15 (1.540), 144.18–20 (1.542), 144.24–5 (1.544), 147.32–3 (1.553)	**618 words**: *ACO* II.1 p. 90.17–91.14 + 92.5–8 + 94.24–95.15 (1.157+ 164+ 185, petition)
Photius of Tyre	**424 words**: *ACO* II.1 p. 291.16–22 (4.9.15), 112.8–10 (4.37), 112.33–6 (4.46), 369.37–9 (9.22), 375.22 (10.10), 377.21–31 (11.22), 462.32–5 (19.3), 464.18–22 (19.10), 465.5–8 (19.18), 465.28–31 (19.24), 465.33 (19.26), 466.17–20 (19.34)	**437 words**: 463.10–464.11 (19.7, petition)

A couple of variables may affect, to a small degree, the homogeneity and reliability of the corpus. First, the samples are taken from records of different gatherings (for example, Eutyches' statements are from the Resident Synod of Constantinople in 448, from the hearings of Constantinople in 449, and from Second Ephesus in 449; Photius' statements are all from Chalcedon); as we have already seen in Section 2, records of different gatherings may not be equally reliable.

Second, while most spoken statements were unprepared, for they arose in the course of the debate, there is evidence that some were prepared beforehand in writing, although the minutes do not mention this. That is the case of a testimony of the presbyter John delivered at the Resident Synod of Constantinople in 448; the Acts introduce this with 'John said', but, at the hearings of the next year, John produced the *aide-memoire* (*hypomnestikon*) on which he had based his testimony, and it appears that he had not deviated much from it. Here are a few lines of John's statement at the Resident Synod of Constantinople (5), followed by the corresponding ones from his aide-memoire (6):

(5) *ACO* II.1 p. 124.4–7 (1.359), also read out at p. 159.5–8 (1.643)

Πρώην τοῦ θεοφιλεστάτου ἐπισκόπου Εὐσεβίου προσελθόντος ἐν συνεδρίωι τῆι ὑμετέραι ἁγιωσύνηι καὶ αἰτιασαμένου τὸν εὐλαβέστατον πρεσβύτερον καὶ ἀρχιμανδρίτην Εὐτυχῆ καὶ βιβλίον ἔγγραφον ἐπιδεδωκότος τὸ καταδηλοῦν νοσεῖν αὐτὸν τὰ τῶν αἱρετικῶν πάθη …

Previously, as the most God-beloved bishop Eusebius appeared at the assembly before your holiness and accused the most devout presbyter and archimandrite Eutyches and had presented a written document declaring that he suffered from the disease of the heretics …

(6) *ACO* II.1 p. 160.34–7 (1.648)

Πρώην τοῦ θεοφιλεστάτου Εὐσεβίου ἐν συνεδρίωι προσελθόντος τῆι ὑμετέραι ἁγιωσύνηι καὶ κατηγορήσαντος Εὐτυχοῦς τοῦ εὐλαβεστάτου πρεσβυτέρου καὶ ἀρχιμανδρίτου καὶ βιβλίον ἐπιδόντος κατ’ αὐτοῦ τὸ καταδηλοῦν νοσεῖν αὐτὸν τὰ τῶν αἱρετικῶν πάθη …

Previously, as the most God-beloved Eusebius appeared at the assembly before your holiness and brought an accusation against Eutyches, the most devout presbyter and archimandrite, and presented a document against him declaring that he suffered from the disease of the heretics …

If we had used John's testimony as a sample of spoken language, that would have led us astray. In this case, we are lucky that his aide-memoire is preserved in the minutes; but it is a fair guess that this was not the only time when somebody spoke using notes and the Acts simply tell us that he gave a speech. When we compare spoken utterances and written documents in the Acts, we must take that into consideration.

Third, we cannot be entirely sure that the written texts presented by each attendee were actually written by them; in case somebody helped them with the composition of the speeches, these would not represent their written language faithfully.

As is clear, there are certain challenges to the use of this material as evidence for spoken Greek. However, I should like to argue that we must not be discouraged, for such challenges are not insurmountable; quite the opposite, the findings of the next Section look very promising.

4.3 Spoken and written language: comparison and findings

Researchers of modern languages have identified several differences between spontaneous spoken language and written language (see Miller and Weinert 1998, with bibliography). In what follows, I am going to investigate whether the

same differences can be observed in the samples of spoken and written language from the Acts of the Council of Chalcedon.

As I have said in Section 4.1, I shall concentrate on aspects of syntactic structure and complexity and on lexicon. In this respect, the differences I find most relevant to our analysis are the following: in spontaneous spoken language a smaller quantity of information is assigned to phrases and clauses than in written language; there is less grammatical subordination and more coordination or parataxis, and the clausal constructions are less complex; the vocabulary is less rich; some constructions that occur in spontaneous spoken language do not occur in written language, and *vice versa* (cf. Miller and Weinert 1998: 22–3).

Based on this account, I shall look at the following elements in the spoken and written samples: the average (and maximal) length of the complex sentence, accepting the editorial punctuation;[52] the number of independent and dependent clauses and the ratio between them; the number of different words used (excluding proper names) and the type/token ratio (TTR), an index of lexical diversity obtained dividing the total number of different words by the total number of words. As for the question of constructions that occur only or preferably in either spoken or written language, I am going to look at participial constructions as competing with finite subordinate clauses for temporal, causal, concessive, final and conditional expressions; I have chosen this type of constructions so as to verify the hypothesis that the participial system underwent a formal and functional reduction in the spoken Koine of the Roman period, to the advantage of finite clauses (see Horrocks 2010: 94, 181–2). To this effect, I am going to find out how many participles are used where finite subordinate clauses may have been used, and *vice versa*; I am also going to calculate the percentage of participial constructions within all such subordinates.

In order to ease the comparison between different samples, in the case of independent/dependent clauses and participial constructions/finite subordinate clauses, I have also normalised frequencies based on the word count of the smallest sample (383 words in Bassianus' written petition) and indicated them between brackets; in the case of the number of different words used, on the other hand, I have simply considered the first 383 words of each sample, excluding proper names. The findings are shown in Table 3 (for each of the five attendees

---

[52] Modern research concludes that 'the sentence is not a useful analytical unit for informal spoken language' (Miller and Weinert 1998: 22); however, it can be used as an analytical unit in the Acts, for the notaries were bound to use the traditional units of written language in taking the records, the Acts being an official written document; it is also to be expected that in a formal occasion such as a council the speakers would have mostly used a formal register.

considered, the left column contains the data from the spoken statements, the right column those from the written statements).

Table 3.	Anatolius		Bassianus		Eusebius		Eutyches		Photius	
	Sp. (632)	Wr. (631)	Sp. (390)	Wr. (383)	Sp. (633)	Wr. (636)	Sp. (618)	Wr. (618)	Sp. (424)	Wr. (437)
Average sentence length (and max.)	31.6 (82)	105.2 (253)	21.7 (47)	34.8 (74)	35.1 (101)	57.8 (133)	17.2 (48)	61.8 (135)	17 (54)	36.4 (118)
Independent clauses (and norm.)	26 (15.7)	13 (7.9)	35 (34.4)	14 (14)	27 (16.3)	14 (8.4)	58 (35.9)	20 (12.4)	36 (32.5)	10 (8.8)
Dependent clauses (and norm.)	50 (30.3)	56 (34)	25 (24.5)	37 (37)	68 (41.1)	61 (36.7)	58 (35.9)	59 (36.6)	33 (29.8)	47 (41.2)
Ratio independent: dependent	1:1.9	1:4.3	1:0.7	1:2.6	1:2.5	1:4.3	1:1	1:2.9	1:0.9	1:4.7
Different words used[53] (and %TTR)	179 (47)	188 (49.1)	146 (30.5)	169 (44.1)	158 (41.2)	176 (45.9)	123 (32.1)	174 (45.4)	157 (41)	178 (46.4)
Participles[54] (and norm.)	9 (5.4)	24 (14.6)	13 (12.8)	13 (13)	15 (9.1)	17 (10.2)	1 (0.6)	23 (14.2)	7 (6.3)	15 (13.1)
Finite subordinates (and norm.)	10 (6.1)	6 (3.6)	2 (2)	4 (4)	3 (1.8)	5 (3)	14 (8.7)	2 (1.2)	3 (2.7)	5 (4.9)
% particip./ sub-ordinates	47.3	80	86.7	76.4	83.4	77.4	6.6	92	69.9	75

As is clear, these figures show that the differences between spoken and written language identified in modern languages can be observed in this material as well:

---

[53] Based on cut-offs of 383 words, corresponding to the size of the smallest sample (Bassianus' petition).

[54] Among participial constructions, conjunct participles (CP) are more frequent than genitive absolutes (GA) in both spoken and written samples. Anatolius: spoken CP 5, GA 4, written CP 19, GA 5; Bassianus: spoken CP 10, GA 3, written CP 11, GA 2; Eusebius: spoken CP 10, GA 5, written CP 14, GA 3; Eutyches: spoken CP 1, GA 0, written CP 17, GA 6; Photius: spoken CP 6, GA 1, written CP 11, GA 4.

In the samples of each attendee the spoken statements have, on average, shorter sentences than the written statements (e.g. 31.6 *vs* 105.2 words in Anatolius).

The spoken statements have many more independent clauses than the written statements, and somewhat fewer dependent clauses (the only exception being Eusebius' samples); this means that for each independent clause there are fewer dependent clauses in the spoken samples, that is, less subordination (e.g. in the case of Anatolius' spoken statements, there are 1.9 dependent clauses for each independent clause, while in his written statements there are 4.3 dependent clauses for each independent clause).

The lexicon is less rich in all spoken samples: for example, the type/token ratio is 47 per cent in Anatolius' spoken statements as opposed to 49.1 per cent in his written statements, and the gap is higher in the samples of the other attendees (e.g. 30.5 per cent *vs* 44.1 per cent in Bassianus' samples).

Of course, there is a degree of variation across samples of different attendees, which may be due to such factors as different communicative situations, register, idiolect and so on. For example, the average sentence is twice as long in Eusebius' spoken samples as in the spoken samples of his arch-enemy Eutyches (thirty-five and seventeen words, respectively); this must be due to the 'oratorial' character of the statements of Eusebius, who was a trained lawyer, as opposed to the brevity of Eutyches' statements, who was answering charges of heresy and trying to give away as little information as possible, while presenting himself as a humble man with little interest in theological subtleties.[55] To be sure, the sentences in Eusebius' spoken statements are as long on average as in Bassianus' and Photius' written samples. Also, the sentences in Anatolius' letter to Pope Leo are much longer than those of the others (105 words on average, but the longest sentence has as many as 253 words), which might reflect the conventions of letter writing as opposed to those of petitions. The syntax of Eusebius' spoken pronouncements is also more complex, having a ratio of dependent clauses to independent clauses of 2.5 to 1, which is almost as high as that of Bassianus' and Eutyches' written samples (2.6 to 1 and 2.9 to 1, respectively).

On the other hand, the data concerning participles are not as straightforward to interpret: participial constructions are generally preferred to finite subordinates to express temporals, causals and so on, and the figures are similar across spoken

---

[55] Cf. Willi (2010: 307–8), who compares specimens of oratory, historiography and 'conversational' literature (from Plato's *Gorgias* and Aristophanes' *Clouds*) and finds that sentences are on average shorter in the last.

and written samples. The only major exception is in Eutyches' samples: he uses only one participle as opposed to seventeen finite subordinates in his spoken statements, while using almost exclusively participial constructions in his written texts (twenty-three as opposed to only two finite subordinates). Anatolius and Photius also use participles less frequently in spoken statements than in written texts, although the gap is not as striking as in Eutyches' samples (nine *vs* twenty-four in Anatolius and seven *vs* fifteen in Photius). The figures show that at this time the participle was still alive in the spoken language of educated people at a formal occasion. But how about the case of Eutyches? We have seen in Section 4.2 that most of his spoken statements are short answers given during a questioning session, as opposed, for example, to the long and 'oratorial' pronouncements of his archenemy Eusebius; this might suggest that, if not in spoken language altogether, in a 'conversational' register such as that of Eutyches' answers, the use of participial constructions was somewhat restricted.[56]

Now that we have ascertained that the spoken language of the individual attendees was different from their written language in a way that matches the modern descriptions of spontaneous spoken language, we can go a little further and attempt to produce generalizations by looking at the same parameters based on the whole corpus of spoken and written samples, without differentiating for different attendees. Here I shall calculate the average sentence length based on the whole corpus, as in Table 3; for the average number of independent and dependent clauses, participial constructions and finite subordinates, I shall use the normalised frequencies given in Table 3; for the number of different words used, I shall consider the sum of the cut-offs considered in Table 3.

The figures in Table 4 confirm the findings of Table 3, while also showing that, on average, participial constructions are more frequent in written than in spoken language.

Table 4.

	Spoken	Written
Average sentence length	23	54.1
Independent clauses (normalised)	27	10.3
Dependent clauses (normalised)	32.2	37.1
Ratio independent/dependent cl.	1:1.2	1:3.6

---

[56] Cf. again Willi (2010: 307–8), who observes a 'more restrictive use of participial phrases' in texts of a 'conversational' character compared to oratory and historiography.

Different words used (and %TTR)	467 (24.4)	588 (30.7)
Participles (normalised)	6.8	12
Finite subordinates (normalised)	4.3	3.3
% participles/subordinates	61.3	78.4

5 Conclusions

In this paper, I have looked at the Acts of the Council of Chalcedon, a unique source for the history and language of the mid-fifth century CE. I have focused on the work of notaries in producing the minutes of the Council, and on how their work has shaped the reliability of the Acts as a historical and linguistic document. I have shown that, as far as history is concerned, the Acts are by and large reliable, while also being affected by some degree of omission and, to a lesser extent, falsification. When it comes to language, my preliminary investigation has shown that the Acts prove precious in pinning down features of the Greek spoken by educated men at this time; if one looks at syntactic complexity and lexicon of spoken statements as opposed to originally written passages, one finds the same differences between spoken and written Greek that have been identified in modern languages: spoken Greek had shorter sentences, less complex clausal constructions and a lesser range of vocabulary than written Greek. The samples also suggest that participial constructions, which are believed to be yielding to finite subordinates in the spoken Greek of this time, were still alive in the use of educated people at formal occasions; at the same time, one sample of a more 'conversational' character shows a restricted use of participial constructions to the advantage of finite subordinates. But this is only the beginning of linguistic research into the conciliar Acts, and I am confident that further investigation into the syntax and lexicon of the minutes will greatly contribute to our appreciation of spoken Greek in the fifth century CE.

References

Akinnaso, F. Niyi 1985. 'On the similarities between spoken and written language', *Language and Speech* 28: 323–59.

Graumann, Thomas 2009. '"Reading" the first Council of Ephesus (431)', in R. Price and M. Whitby (eds.), pp. 27–44.

Graumann, Thomas 2018. 'Documents, acts and archival habits in early Christian Church councils: a case study', in A. Bausi, Ch. Brockmann, M. Friedrich, and S. Kienitz, (eds.), *Manuscripts and Archives. Comparative Views on Record-Keeping*. Berlin – Boston: De Gruyter, pp. 273–94.

Horrocks, Geoffrey 2010. *Greek: A History of the Language and its Speakers*, 2nd edn. Chichester: Wyley-Blackwell.

Koch, Peter 1995. 'Une langue comme toutes les autres : latin vulgaire et traits universels de l'oral', in L. Callebat (ed.), *Latin vulgaire – latin tardif* IV. Hildesheim/Zürich/New York, pp. 125–44.

Koch, Peter and Oesterreicher, Wulf 2011. *Gesprochene Sprache in der Romania*, 2nd edn. Berlin – New York: De Gruyter.

Lancel, Serge 1972. *Actes de la conference de Carthage*. Paris, Les éditions du Cerf.

Mari, Tommaso 2018. 'The Latin translations of the Acts of the Council of Chalcedon', *GRBS* 48: 126–55.

Millar, Fergus 2006. *A Greek Roman Empire. Power and Belief under Theodosius II (408/450)*. Berkeley – London: University of California Press.

Miller, Jim and Weinert, Regina 1998. *Spontaneous Spoken Language. Syntax and Discourse*. Oxford: Clarendon Press.

Oesterreicher, Wulf 1997. 'Types of orality in text', in E. Bakker and A. Kahane (eds.), *Written Voices, Spoken Signs. Tradition, Performance, and the Epic Text*. Cambridge (Mass.) – London: Harvard University Press, pp. 190–214.

Price, Richard 2009. 'Truth, omission, and fiction in the Acts of Chalcedon', in R. Price and M. Whitby (eds.), pp. 92–106.

Price, Richard, Booth, Phil and Cubitt, Catherine 2014. *The Acts of the Lateran Synod of 649*. Liverpool University Press.

Price, Richard and Gaddis, Michael 2005. *The Acts of the Council of Chalcedon*, 3 vols. Liverpool University Press.

Price, Richard and Whitby, Michael (eds.) 2009. *Chalcedon in Context. Church Councils 400-700*. Liverpool University Press.

Riedinger, Rudolf 1982. 'Die Lateransynode von 649 und Maximos der Bekenner', in F. Heinzer and Ch. Schönborn (eds.), *Maximus Confessor: Actes*

du Symposium sur Maxime le Confesseur, Freiburg/Schweiz 2.-5. September. Freiburg, Éditions Universitaires, pp. 111–21.

Roueché, Charlotte 2009. 'Acclamations at the Council of Chalcedon', in R. Price and M. Whitby (eds.), pp. 169–77.

Schwartz, Eduard 1929. *Der Prozess des Eutyches*. Munich, Sitzungsberichte der Bayerischen Akademie der Wissenschaften, Philosophisch-historische Klasse, Abt. 5.

Schwartz, Eduard 1933. 'Zweisprachigkeit in den Konzilsakten', *Philologus* 88: 245–53.

Schwartz, Eduard 1933–7. *Acta conciliorum oecumenicorum. Tomus II: Concilium Universale Chalcedonense*, 4 vols. Berlin – Leipzig: Walter De Gruyter & Co.

de Ste. Croix, Geoffrey Ernest Maurice 1984. 'A worm's-eye view of the Greeks and Romans and how they spoke: martyr-acts, fables, parables and other texts', *Latin teaching* 37.4: 16–30.

de Ste. Croix, Geoffrey Ernest Maurice 2006. *Christian Persecution, Martyrdom, & Orthodoxy*, Oxford University Press.

Teitler, Hans C. 1985. *Notarii and Exceptores. An Inquiry into Role and Significance of Shorthand Writers in the Imperial and Ecclesiastical Bureaucracy of the Roman Empire (from the Early Principate to c. 450 A.D.)*. Amsterdam: Gieben.

Vatri, Alessandro 2017. *Orality and Performance in Classical Attic Prose: A Linguistic Approach*. Oxford University Press.

Verano, Rodrigo 2018. 'The truth alone will suffice: traces of spoken language in Plato's *Apology of Socrates*', *Scripta Classica Israelica* 37: 25–43.

Weidmann, Clemens 2018. *Collatio Carthaginensis anni 411*. Berlin – Boston: De Gruyter.

Wessel, Susan 2001. 'Literary forgery and the monothelete controversy: some scrupulous uses of deception', *GRBS* 42: 201–20.

Whitby, Michael 2009. 'An unholy crew? Bishops behaving badly at church councils', in R. Price and M. Whitby (eds.), pp. 178–96.

Willi, Andreas 2010. 'Register variation', in Bakker, Egbert J. (ed.), *A Companion to the Ancient Greek Language*. Oxford: Wyley-Blackwell, pp. 297–310.

8. Bilingual Letter Writers: The Verbs γράφω, οἶδα and θαυμάζω in Formulae, Idioms and Collocations

Victoria Beatrix Fendel

1 Introduction

Any verb is surrounded by an argument structure and a participant structure, which map onto each other. Thus, there is a morphosyntactic dimension (the c-structure in Lexical-Functional Grammar (LFG) terms) and a semantic dimension (the f-structure in LFG terms) (Bresnan, Asudeh, Toivonen and Wechsler 2015). Many verbs allow for more than one pattern. If the same lexical verb appears with more than one pattern, not only the morphosyntax but also the semantics of the verb change. Such is the case for γράφω, οἶδα and θαυμάζω, in a Greek-Coptic corpus of private letters on papyrus dating from the early Byzantine period and originating from Middle and Southern Egypt.

The range of argument and participant structures that a lexical verb can appear in has been captured in the form of verb profiles. Profiles have been developed for French verbs in the context of the Lexicon-Grammar Approach (LADL) (Gross 1984) and for German verbs in the context of the Leipzig ValPal project (www.valpal.info; similarly for English, see Hanks (1996, 2013)). Profiles of (seemingly) semantically equivalent verbs across languages often differ. Compare for instance 'to listen to', 'etw. hören' and 'écouter qqch'. While the German and French constructions include a second argument referring to what is being listened to without a preceding preposition, the English verb calls for a preposition preceding the argument referring to what is listened to. Grossman (2019) shows for Greek and Coptic specifically that in the case of loan verbs, the argument structure of the model language may not be preserved in the replica language[1] but that the verb in question may be fit into the pattern of the equivalent native verb.

In the early Byzantine period, Egypt had been a bilingual region for more than a millennium. Language contact had initially been limited to the trade *metropoleis* of the north (Bergeron 2015; Torallas Tovar 2010: 255; Villing 2015), but spread southwards during the Ptolemaic and Roman periods (Adams 2003:

---

[1] For the terminology, see Matras (2009).

534; Crespo 2007; Fischer-Bovet 2014: 23; Hall 2014: 268; Kraus 2000). Over the course of the early Byzantine period, the political situation and with it the relative status of Greek and Egyptian, called Coptic from the fourth century onwards, changed.

Coptic advanced into the official sphere, a former stronghold of Greek. Three good examples of this are (i) the translation of the New Testament into Coptic with the increasing importance of Christian faith, as early as the second century (Wright and Ricchuiti 2011: 497) (ii) the emergence of the Coptic alphabet from around AD 100 onwards (Fendel 2021; Quack 2017), (iii) the use of Coptic in business communications, as early as the Douch ostraca in the fourth century (Choat 2009: 347) and eventually wills, such as the well-known wills from the monastery of St Phoibammon in the early seventh century (Fournet 2019; Krause 1969). Thus, in the cultural, social and political spheres Coptic was gaining in importance and achieving an almost equal status to Greek over the course of time.

From a linguistic perspective, the impact of Egyptian on Greek has been explored for Ptolemaic collections, such as the *agoranomos* contracts (Vierros 2007, 2012), texts written with a rush (Clarysse 1993), and the archive of Kleon and Theodoros (Clarysse 2010), as well as Roman collections, such as the Narmouthis ostraca (Bagnall 2007; Leiwo 2003; Rutherford 2010). It seems that at least from the Roman period onwards, a regional variety of Greek in Egypt had developed and thus must be taken into account (Dahlgren 2016, 2017).[2] By contrast, the early Byzantine period has often been passed over[3] potentially because of it being a transitional period historically speaking (Keenan 2007; Kiss 2007; van Minnen 2007) or because comparative literary sources are scarce (e.g. the New Testament for the Roman period).

What the present chapter is not is a comprehensive overview of language contact in late antique Egypt; research from the historical, socio-historical and linguistic perspectives has been done. Furthermore, this chapter is based on a specific corpus of texts. Thus, the results presented here apply to this very corpus. In order to apply them to different corpora, they would first have to be verified for these. In addition, this chapter is not a study of Coptic in the first place, but a

---

[2] The impact of Greek on Egyptian is not our primary concern here, but has also been studied (Grossman, Dils, Richter and Schenkel 2017; Hasznos 2006, 2012).

[3] Volumes on language contact and bilingualism in Egypt pass over the early Byzantine period, e.g. Cromwell and Grossman (2017) (with chapter 11 on texts dating up to the third century and chapter 12 on texts dating from the eighth century onwards).

study of Greek; research on the traces which the contact situation between Greek and Coptic has left on the Coptic side is relevant only insofar as it proves that there was a contact situation that worked in both directions; research on Coptic linguistics is relevant to the extent that Coptic and Greek interacted and we need to know what the interactant Coptic looked like in order to recognise its impact on Greek. Finally, the approach taken here is descriptive and linked to the LFG framework; no conclusions about typological issues are claimed.

The aim of the present chapter is to propose a method to approach verbal complementation patterns, that is the collocation-idiom-formula continuum. This method applies the principles of the variationist paradigm to verbal complementation patterns, that is the distinction between standard, variation and deviation. The chapter takes a primarily theoretical approach.

The chapter is divided into four Sections. Section 2 introduces the reader to the method of conceptualising verbal syntax that is applied here. Section 3 considers verbal complementation patterns from a typological point of view as well as Greek and Coptic verbal complementation patterns in particular. Section 4 introduces the corpus of texts and applies the approach introduced in Section 3 to the corpus data. Section 5 summarises the results and concludes by evaluating the usefulness of the approach suggested here.

2 Conceptualising verbal syntax in a language-contact setting

We define verbal syntax as the structure surrounding a lexical verb or verbal phrase (in the case of verbal Multi-Word Expressions). At the morphosyntactic level, this is the argument structure that the verb subcategorises for. At the semantic level, the participant structure that the verb subcategorises for maps onto the argument structure. When analysing data, the structure surrounding a lexical verb or verbal phrase may or may not comply with the expected standard in a language. In the latter case, we must ask what caused the difference between the expected standard and the observed structure. One of the potential reasons, a reason that is unique to language-contact settings, is the adoption of a structure from another language with or without adaptation of this structure.

With Matras (2009: 238) we call the language that adopts or receives a structure the replica language and the language that borrows or gives a structure the model language. Interaction between two languages can happen at several levels. In order to distinguish these, we make a tri-partite distinction reminiscent of Myers-Scotton's (2002) Abstract-Level Model and the two-tier

approach of LFG. We distinguish a surface level, that is the morphosyntactic level (the c-structure in LFG), an underlying structural level, that is level of the grammatical and semantic relations (the f-structure in LFG), and a deeper conceptual level, which captures culturally determined differences. This third level is relevant for example when discussing formulae in different languages (e.g. epistolary formulae). These have usually evolved out of a specific cultural setting and tradition and thus differ between languages, e.g. 'to write' vs. 'to say' in Greek and Coptic epistolary formulae (Depauw 1997; Choat 2007, 2010; Koskenniemi 1956).

Differences between the expected standard and an observed structure may be contextually motivated. We therefore distinguish three types of contexts, that is formulaic, semi-formulaic and free contexts. The select corpus of texts consists of private letters. In these, the formulaic sections appear at the beginning and the end, comparable to English 'dear X' and 'best wishes' (e.g. Fournet 2009; Kim 2011; Koskenniemi 1956; Stolk and Nachtergaele 2016); the semi-formulaic sections appear either close to the beginning and end or they are interspersed in the letter body; the free section of a letter is the letter body. Semi-formulaic expressions are pragmaticalised expressions that structure the letter body as signposts and / or hedges, such as the disclosure formula βούλομαί σε γνῶναι ὅτι 'I want you to know that' (Davis and Maclagan 2020; Erman 2001; Hulleberg Johansen 2021; Porter and Pitts 2013).

We distinguish these three types of contexts firstly because in formulaic contexts, a minor variation may disrupt the formulaic structure completely, whereas in freer contexts, minor variations may even be overlooked at times. Secondly, research on speech production as well as research on language learning suggest that speakers / writers and learners approach formulaic contexts differently from less fixed contexts (e.g. Namba 2010; Weinert 2010; Wood 2010; Wray 2009). Thus, we expect there to be differences in language usage between the three types of contexts and we expect these differences to manifest in the types and frequency of the nonstandard structures that appear.

Importantly, not every structure that differs from the expected standard is a language-contact phenomenon. First of all, the expected standard is difficult to determine for early Byzantine Greek in Egypt not only because we are lacking a comprehensive description of the language (cf. Gignac's (1976) third volume), but also because a wide range of parameters has to be taken into account when considering our texts. Greek had developed over the centuries, yet not always in a linear manner (modernisms), and at the same time backward-looking movements (classicisms) had occurred (e.g. Adams 2013; Lee 2013; Luiselli

1999). As mentioned, at least from the Roman period onwards, a regional variety of Greek in Egypt seems to have existed (regionalisms). Our texts originate from different situational contexts (register-related variation), and from different societal groups (sociolectal features) as well as from a range of individuals (idiolectal features). Secondly, the preservation of texts is in essence by chance. Thus, we may have an incomplete view of early Byzantine Greek. This calls for caution when drawing inferences. Thirdly, even language learners do not constantly draw on their first language or any other language they know, but they are often influenced by learning methods and by the context of learning and thus produce structures that are nonstandard but not clearly due to drawing on another language (Birney et al. 2019; Cheng et al. 2021; Evans 2012a, 2012b).

Distinguishing between types of nonstandard structures is often complicated. Therefore, we distinguish broadly between features that differ from the expected standard but are grammatically correct and those that differ from the expected standard and are grammatically incorrect. We call the former variations and the latter deviations. Most variations are context-dependent (register-related),[4] whereas most deviations are idiolectal (cf. Labov's (1991) variationist paradigm). However, given our incomplete vision of what the early Byzantine Greek standard was, some of our deviations may in fact be patterns that were part of the standard repertoire. Equally, there may be variations that we can plausibly explain, but that were in fact not part of the standard repertoire. This is an issue which every corpus language is faced with and that cannot be completely resolved, but that one has to bear in mind.

Finally, we mentioned that not every deviation is caused by the interaction of two languages. Equally, not every deviation that is caused by the interaction of two languages is a case of bilingual interference. In fact, there are two main options, that is bilingual interference and convergence. Bilingual interference is idiolectal and either momentary (one-offs) or temporary (interlanguage) (Matras 2009: 74–79 and 310–312; Myers-Scotton and Jake 2000; Selinker 1972). Interferences are hence comparatively rare and are not spread across writers (or only by chance), they are ungrammatical[5] and they show no or a low degree of adaptation of the model structure to the replica language. Conversely, convergence is a gradual

---

[4] Variations can be variations towards more elaborate or classicising patterns as well as variations towards simpler or colloquial patterns.

[5] We can identify in the papyrological data only instances of negative transfer, whereas those of positive transfer may go unnoticed (Butler and Hakuta 2004: 129–34).

process by which two languages merge. The structure that is incorporated into the replica language is adapted to the system of the replica language so as to be grammatical in the replica language. In order to distinguish between instances of interference and convergence, we consider whether an observed structure is grammatical (and idiomatic, in the case of formulaic sections), is frequent and spread across writers and texts, and to what extent the structure is adapted to the replica language, here Greek.

3 Verbal complementation patterns

3.1 Types of complements

The argument slot(s) of a verb can in many cases be filled by several types of linguistic material. We are not interested in the subject slot and call all the other arguments that a verb subcategorises for complements. In essence, complements can be nominal or verbal and phrasal or clausal. Compare for instance *I saw the house* with the two-argument verb 'to see' and a phrasal argument in the second argument slot, that is *the house*, a Noun Phrase (NP), with *I saw that the house was damaged* with the same two-argument verb yet this time with a clausal argument in the second argument slot, that is *that the house was damaged*. Between these extremes, an NP and a clause, there is a range of verbal and nominal phrases and clauses that can fill the complement slot. To use the same two-argument verb as before and give only a few examples, consider *I saw the house being painted green* with a gerund construction and *I saw the house collapse* with an infinitive construction. In both cases, we have a Verb Phrase (VP) in the argument slot. Lehmann (1988, Section 3.1.4) maps the kinds of complements that can appear in the complement slot on a continuum ranging from nominality to sentiality. From a purely structural perspective, we adopt his idea.

Given that there is often more than one morphosyntactic option for the argument slot of a verb, it has been debated whether choosing one or the other complementation pattern is an analytic process. To put it differently, are verbal complementation patterns semantically compositional or non-compositional? In the former case, we would say that a complementation pattern has an intrinsic meaning and that this meaning is added to the meaning of the lexical verb when the complementation pattern is selected. The meaning of the combination of the verb and the complementation pattern is then the summation of two semantic components. In the latter case, that is taking complementation patterns as

semantically non-compositional, we would say that a complementation pattern does not have one meaning that is intrinsic to it, but that the verb and the complementation pattern interact in some (!) way. Consequently, the semantics of the combination of the verb and the complementation pattern are non-compositional.

Based on their distribution across verbs and the meanings associated with the combinations of a verb and a complementation pattern, Cristofaro (2008: esp. 587–95) argues that Greek verbal complementation pattern are non-compositional.[6] In order to refer to patterns and describe them, we distinguish between finite and infinite patterns on the morphosyntactic level and between factive and prospective patterns on the semantic level with factive to be taken as purely contextually anaphoric rather than in the context of truth-conditional semantics (Schulz 2003). There is no one-to-one mapping between these syntactic and semantic categories as Cristofaro (2008) explains in detail.[7]

3.2 Classes of verbs

So as to put order into our analysis, we organise verbs in classes. The superordinate classes are based on the argument structure and the type of complement that fills a verb's complement slot. The superordinate classes that are relevant to the select corpus of texts are: transitive verbs, intransitive verbs (no direct object!), copular verbs, modal verbs (grammatical relations and semantic contribution), auxiliary verbs (grammatical relations), impersonal verbs (third-person singular subject). Often, these larger classes of verbs fall into smaller semantic groups. These are defined by means of their participant structure.

For example, we have a superordinate class of verbs that subcategorises for a subject and a direct object. Subordinate classes distinguish between complements being nominal or verbal. Further subordinate classes distinguish between the participant roles the complement takes on. Three examples are given in Table 1.

---

[6] However, the way people learn(ed) complementation patterns may have been one in which they were treated as being compositional perhaps in order to simplify at early stages of language learning (see also Tomasello 2003).

[7] Several patterns reflecting degrees of subjectivity, e.g. in result clauses and factive complement clauses, seem to have traded their semantic distinctness for a register-related one in the post-classical period.

Table 1. Classes of verbs.

(a) {agent}[nominative] {theme}[accusative] → speech acts
→ λέγω 'to say'
(b) {agent}[nominative] {patient}[accusative] → action
→ ἀποδιώκω 'to chase away'
(c) {experiencer}[nominative] {stimulus}[accusative] → act of perception
→ βλέπω 'to see / look at'

In [1(a)] to [1(c)], the subject is inflected in the nominative case and the direct object in the accusative case. The combination of participant roles sets the three groups apart. Furthermore, verbs like λέγω *legō* in [1(a)] can alternatively be complemented by an infinitival structure or a factive complement clause. Verbs like βλέπω in [1(c)] can alternatively be complemented by an infinitival or participial structure or a factive complement clause. By contrast, for verbs like ἀποδιώκω in [1(b)], no alternative pattern appears.

Importantly, while there is a valid methodological approach to group verbs into these classes, these classes were established for the select corpus of texts. For a different corpus, the classes of verbs may have to be adapted. For example, in the corpus, there is a clear divide in patterns between verbs of begging, such as παρακαλέω and ἀξιόω, and verbs of command, such as κελεύω and ἐντέλλω. Verbs of begging are primarily complemented with a prospective complement clause; verbs of command preserve infinitival patterns. Yet, verbs of command appear with prospective complement clauses in other corpora, e.g. P.Cair.Zen. 4.59546, 2 (official letter, 257 BC, Philadelphia) ἐνε]τείλατο καὶ σοί, ἵνα κατασκ[ευ]ασθῆι 'he ordered you that it be prepared'.

3.3 Collocation, idiom, formula

Like the French and German verbs mentioned above, many Greek verbs enter into more than one pattern, that is into more than one combination of an argument structure with a participant structure. The organisation of verbs into groups based on their argument and participant structures is carried out empirically. To put it differently, when a verb appears in pattern a, the relevant instance is assigned to group A; when the same verb appears in pattern b, the relevant instance is assigned to group B and so on. Thus, some verbs are assigned to more than one group of verbs.

In order to conceptualise this situation, we view complementation patterns on a continuum ranging from the most basic pattern, the collocation, to the most specific pattern, the formula. The terms collocation, idiom, and formula have received a number of definitions in research literature. For our purposes, they refer to three distinct complementation patterns of the same lexical verb.

For reasons of illustration, we use οἶδα *oida*, a high-frequency item in the select corpus of texts. οἶδα appear in several syntactically distinct patterns, which are linked by regular internal processes at the syntactic (e.g. intransitivisation), pragmatic (e.g. pragmaticalization) and semantic levels (e.g. metaphorical extension). In (1), οἶδα *oida* is combined first with a nominal direct object and subsequently with a factive complement clause.

(1) P.Neph. 18, 15–17
καίτοι οἴδατε τὴν τιμὴν τοῦ σίτου ὅτι γυνή εἰμι, οὐ δύναμαι ἀγοράσαι.

kaitoi	*oid-ate*		*tēn*	*timēn*
yet	know(PERF)-1PL.IND.ACT		the.ACC.SG.F	price-ACC.SG.F

tou	*sitou*	*ʰoti*	*gunē*
the.GEN.SG.N	grain-GEN.SG.M	that	woman-NOM.SG.F

ei-mi,	*ou*	*duna-mai*	*agora-sai*
be(PR)-1SG.IND.ACT	not	can(PR)-1SG.IND.MID	buy(AOR)-INF.ACT

'yet, you know the price of grain (and) that I am a woman (and that) I can (hence) not make purchases'

With both these patterns, οἶδα *oida* falls into the class of verbs of perception (see Section 3.2).[8] This is the pattern in which the verb appears most frequently. The pattern is semantically compositional and syntactically transparent. No contextual constraints apply. We call this basic pattern of a verb its collocation.

---

[8] The classical distinction between the participial pattern to emphasise the act of perception vis-à-vis the ὅτι-pattern to emphasise the factivity of the event observed is retreating. Moreover, οἶδα was used with the ὅτι-pattern, then emphasizing the factivity of the event observed, already in classical literature (e.g. Lysias 1, Sections 1, 22, 28, and 36).

In (2), οἶδα is treated as an intransitive verb through object deletion. It appears in a comparative clause that is pragmaticalised as a hedge[9] for claims (see also la Roi 2022).[10]

2) P.Oxy. 34.2727, 9–11
ὡς γὰρ οἶδας, ἀπαραίτητός ἐστιν ἡ χρεία·

ʰōs gar	*oid-as*		*aparaitēt-os*
as for	know(PERF)-2SG.IND.ACT		irresistible-NOM.SG.M/F

es-tin		*ʰē*	*kʰrei-a*
be(PR)-3SG.IND.ACT		the.NOM.SG.F	need-NOM.SG.F

'For, as you know, the need is irresistible.'

In declarative clauses / claims, responsibility for what is said lies with the speaker (or here writer) (Verstraete 2007: 106). By means of the hedge, the writer limits his responsibility for the propositional content and appeals to the hearer (Hulleberg Johansen 2021: 86). In O.Claud. 4.896, 5–7 an asyndetic option instead of the comparative clause seems to be selected.[11]

Other regular operations, apart from transitivity-related operations as in (2), are the co-existence of classical and post-classical patterns (e.g. ἀξιόω *axioō* 'to deem somebody/something worth of something' vs 'to beg'), factive and prospective patterns (e.g. λέγω 'to tell somebody that something is the case' vs 'to tell somebody to do something'), or grammaticalised and non-grammaticalised patterns (e.g. θέλω 'to want' vs 'will') for the same verb. We call such an alternative pattern that a verb appears in its idiom. This pattern is usually, but not always,

---

[9] A hedge is a lexical or structural element that adds tentativeness to an expression (Hulleberg Johansen 2021: 82), such as English *sort of*, *more or less* and parenthetical *I mean* (Crystal 2008: 227).

[10] In the corpus, the verb λέγω rather than οἶδα seems preferred in these hedges (e.g. PSI 8.938, 5; PSI 8.939, 5; P.Cair.Masp. 1.67061, 4–5). The function is similar, except that the responsibility is shifted to a third party rather than the interlocutor due to the third-person verb forms.

[11] Proximity searches in the *Duke Database of Documentary Papyri* show that ὡς οἶδας is the preferred form for the hedge. ὡς οἶδες only appears in P.Fouad. 85 (AD 501–700), ὡς οἶσθα in P.Oxy. 79.5210 (AD 298–299), and ὡς οἴδατε in P.Oxy. 36.2788 (AD 201–300) and P.Tebt.2.420 (AD 201–300). The hedge appears occasionally already in classical literature, e.g. Isaeus 2, Section 12.

less frequent than the collocation and semantically different. The idiom does not have to be limited in its contextual applicability.

Contextual applicability is, however, limited with the formula of a verb, in that the formula appears only in one or a specific number of fixed contexts. In (3), οἶδα *oida* appears in the so-called disclosure-formula (Porter and Pitts 2013).[12]

> (3) P. Oxy. 16.1830, 15–17
> καὶ ἵν᾽δέναι ἔχυ, ἐν τῇ ὑμετέρᾳ εὐδοκιμησιν ἀναφέρω τὰ πάντα.
>
> | *kai* | *in'* | *d-enai* | *ekʰ-u* |
> |-------|-------|----------|---------|
> | and | in.order.that | know(PERF)-INF.ACT | can(PR)-3SG.OPT.ACT |
>
> | *en* | *tē* | *ʰumeter-a* | *eudokimēs-in* |
> |------|------|-------------|----------------|
> | in | the.DAT.SG.F | your- DAT.SG.F | good.repute-DAT.SG.F |
>
> | *anapʰer-ō* | *ta* | *pant-a* |
> |-------------|------|---------|
> | bring(PR)-1SG.IND.ACT | the.ACC.PL.N | everything-ACC.PL.N |
>
> 'and it (sc. your good repute) may know (that) I will bring everything to your good repute'

The disclosure-formula consists of a deontic form or periphrasis of a verb of learning / realising (e.g. οἶδα) and a factive complement clause containing the piece of (new) information to be flagged. In (3), the writer opts for an asyndetic pattern instead of the regular factive complement clause. We call a verb's formula a pattern which semantically and syntactically differs from a verb's collocation and idiom and which appears in a formulaic context with a very specific function.

We locate collocations, idioms and formula on a continuum ranging from the most basic to the most specific. The collocation, idiom and formula of a verb are each a group of patterns that share a form-function mapping, as is shown in (1) and (3), in that the patterns shown are alternatives to the expected pattern, and in (2), in that an asyndetic pattern appears instead of a complement clause.

---

[12] The disclosure formula appears rarely already in classical literature (e.g. Xenophon, *Anabasis*, 1.3.15) and relatively commonly in the New Testament. In the papyri, it seems to be used across registers and serves to flag new information. In the corpus, the disclosure formula appears more frequently with μανθάνω (9 instances) and γιγνώσκω (11 instances).

3.4 Greek and Coptic

Three aspects in which Greek and Coptic typologically differ are relevant to verbal complementation patterns. On the surface level, it is the aspect of word order; on the structural level, it is the aspects of rectional and combinatory elements with the verb and of the morphosyntactic encoding of complements. We discuss these aspects in turn.

To begin with the word order, Celano (2013b, 2013a) argues that Classical Greek word order is determined by information-structural considerations. However, there are aspects of word order that are more syntactically determined, such as the word order with prepositives and postpositives and the word order inside an NP. A similar opinion is held with regard to later Greek, New Testament Greek, by Kirk (2012). Given that Greek is an inflecting language, the word order is not a primary means for indicating syntactic functions. By contrast, Coptic is a noninflecting language and word order is one of the tools to encode syntactic functions. This is not to say that information structure is not important (e.g. Zakrzewska 2006 on Bohairic). Yet often a different word order pattern has to be selected when information-structural considerations require for instance postponing of an element (Layton 2011). Coptic encodes syntactic functions not only by means of position but also by means of prepositions and incorporation. In this context, Grossman (2015) found that the latter mechanisms cannot be applied before the verb phrase.

To move on to rectional and combinatory elements with the verb, rectional elements are elements that have a solely syntactic function, whereas combinatory elements have first and foremost a semantic function. Some elements may be rectional and combinatory at the same time (Layton 2011: § 181). For example, in ϢⲒⲚⲈ ⲉ-/ⲉⲣⲟ= *šine e-/ero=* 'to visit (somebody)', the entity that is visited is referred to by the complement of ⲉ-/ⲉⲣⲟ= *e-/ero=*. Leaving out ⲉ-/ⲉⲣⲟ= *e-/ero=* would trigger a semantic change in the VP. Thus, ⲉ-/ⲉⲣⲟ= *e-/ero=* in this combination is rectional and combinatory. Since we are primarily interested in the syntax of the verb, we leave combinatory elements largely aside and instead focus on rectional elements. The most important rectional elements in Coptic are Direct Object Marking (DOM) elements, a category of elements that Greek does not have. Most often, the DOM element is ⲛ-/ⲙⲙⲟ= *n-/mmo=*. DOM is obligatory in the durative construction (present tense) with very few exceptions (Grossman 2009 for one hypothesis of how the *status quo* came about; Layton 2011: § 171), but optional in other constructions (Engsheden 2008 on Sahidic; Engsheden 2018 on Lycopolitan).

By contrast, Greek does not operate with rectional elements with the verb. However, one could analyse the repetition of the preverb in order to encode the semantic argument or adjunct of a compound verb, such as ἀπό in Luke 5:2 ἀπ᾽αὐτῶν ἀποβάντες 'those going away from them', as an instance of a rectional element or perhaps rather an element that is rectional and combinatory (Luraghi 2003 on preverbs; Robertson 1919: 557–65) with the verb. This is a phenomenon that appears already in classical texts but seems to gain in frequency in the post-classical period perhaps due to the restructuring of the case system (Bortone 2010). In essence, adverbial case functions were losing ground and were replaced by analytic phrases built with prepositions and postpositions.

Finally, regarding the encoding of complements, Section 3.1. introduced (i) the morphosyntactic distinction between finite and infinite patterns, (ii) the semantic distinction between factive and prospective patterns, and (iii) Lehmann's continuum ranging from nominality to sentiality as regards the type of complement. Regarding (i), Bentein (2017) argues for Greek that the post-classical tendency towards more finite and fewer infinite complements correlates with a pragmatic restructuring of the complementation system. In essence, finite complementation patterns became the standard option with infinite ones being available primarily in high-register discourse. Regarding (ii), Greek operates a two-tier system where factive complement clauses are usually headed by ὅτι or ὡς (James 2008 for the register-related difference) and prospective ones by ἵνα or ὅπως (Clarysse 2010; Hult 1990 for the register-related difference). Regarding (iii), it should be noted that formally asyndetic structures that rely on logical subordination lie strictly speaking outside the continuum. Asyndetic patterns differ not only syntactically, but also semantically and / or pragmatically from their syndetic counterparts (Debaisieux 2004; Gries and Stefanowitsch 2004; Penner and Bader 1995: 96–117; Ruiz Yamuza 2020).

By contrast, Hasznos (2012, 2017) finds a preference for infinite complementation patterns with verbs of request (begging) across Coptic dialects. Interestingly, however, Greek loan words are more prone to finite complementation patterns (e.g. ⲡⲁⲣⲁⲕⲁⲗⲉⲓ ⲭⲉⲕⲁⲥ *parakalei ďekas* 'to beg that'). Finite complement clauses, both semantically factive and semantically prospective ones, are mostly headed by ⲭⲉ *ďe* (see also Müller 2012). This contrasts with the two-tier system of Greek mentioned above. Coptic does not have morphosyntactically distinct patterns for dependent and independent clauses. In this context, Shisha-Halevy's (2007) zero-conjunction constructions in Bohairic Coptic are notable. These are equivalent to English *I saw the house was damaged*, where *the house was damaged* fills the complement slot of *see* but no complementiser is present. We call these

constructions asyndetic. They exist in the Greek and Coptic parts of the select corpus of texts but are not especially frequent. As mentioned, they lie strictly speaking outside the sphere of the structural complementation of verbs.

4 Operationalising the collocation-idiom-formula approach

The select corpus of texts consists of all private letters belonging to bilingual (Greek-Coptic) papyrus archives (collections of texts) dating from the early Byzantine period, that is the fourth to mid-seventh centuries. The relevant archives are: Apa John, Apa Nepheros, Apa Paieous, Papyri from Kellis, Dioscoros of Aphrodito (see further www.trismegistos.org). There are 127 Greek and 137 Coptic letters in total. The number of Greek words in the corpus is 13,609. The number of VPs in the corpus is 2,045, our total of tokens, and the number of distinct lexical verbs in the corpus is 436, our total of types. The three types chosen here, that is γράφω, οἶδα and θαυμάζω, account for 91 tokens, that is about 4 per cent of the total of tokens.

We apply the same methodology applied to οἶδα above to γράφω and θαυμάζω. The collocation of γράφω is a verb of communication, that is an indirect object referring to a recipient (in the dative case) and a direct object referring to a message (in the form of a factive complement clause); the idiom of γράφω is a verb of command, that is an indirect object referring to a recipient (in the dative case) and a direct object referring to an order (in the form of a prospective infinitival structure); the formula of γράφω is a verb of request, that is an indirect object referring to a recipient (in the dative case) and a direct object referring to a request (in the form of a prospective complement clause) (Fendel 2022a).

θαυμάζω has only a collocation and a formula. The collocation of θαυμάζω is a verb of emotion, that is an intransitive verb expanded by a causal clause providing a reason for the emotion; the formula of θαυμάζω is the topic-shift formula (Fendel 2022b, chap. 9), as in P.Kell. 1.65, 3–5 θαυμά[ζ]ω πῶς οὐκ ἔγραψάς μοι μίαν ἐπιστολὴν περὶ οὐδενὸς ἁπλῶς 'I am wondering why you did not write to me any letter about anything at all'. θαυμάζω appears in the first person singular; the indirect question following contains a verb in a second-person past tense usually referring to a past failure on the addressee's part.

After assigning tokens to groups and establishing profiles of verbs, we apply the distinction between variations and deviations introduced in Section 2 by taking into considerations the range of factors that impact on our writers' use of

language. Table 3 summarises the distribution of variations and deviations in the select corpus of texts.

Table 3. The syntax of γράφω, οἶδα oida, and θαυμάζω.

		Standard	**Variation**[13]	**Deviation**	**TOTAL**
γράφω	C	35	ø	2[14]	37
	I	4	4	ø	8
	F	6	4	3[15]	13
	lost				3
θαυμάζω	C	1	ø	ø	1
	F	5	1	2[16]	8
οἶδα	C	8	2	ø	10
	I	1	7	ø	8
	F	2	ø	ø	2
	lost				1
TOTAL		62	18	7	91

When we apply this distinction to the complementation patterns of our three select verbs, we notice that variations appear with all three verbs and with almost every type of complementation pattern, whereas deviations are limited to γράφω (collocation and formula) and θαυμάζω (formula).

The method of establishing verb profiles suggested here is multi-layered, in that we establish groups of verbs based on c-structures, sub-groups within these based on f-structures, and subdivisions within these based on the production circumstances. The latter are relevant to the select corpus, since it consists of documentary texts. The verb profiles established allow for comparison and contextualisation of (i) variant and deviant patterns, (ii) modernised and restructured patterns, and (iii) the linking of patterns of the same verb which have arisen from regular internal e.g. transitivity-related operations or metaphorical

---

[13] Variation downwards includes asyndetic and paratactic structures; variation upwards includes infinite (participial and infinitival) structures and circumstantial participles inserted for clarification of the intended pattern of the verb (e.g. P.Lond. 6.1917, 23 ἐντελλώμαινοι added to γράφω to indicate that the idiom is intended).

[14] P.Herm. 7, 6 and P.Kell. 1.5, 21, see Fendel 2022b.

[15] P.Lond. 6.1916, 33, see Fendel 2022b; P.Kell. 1.65, 5 and 31, see Fendel 2022a.

[16] P.Kell. 1.68, 10–11 and P.Kell. 1.64, 5–6, see Fendel 2022b.

extension (see e.g. Jiménez López 2016). Moreover, for a corpus that is set in a language-contact situation, the establishment of verb profiles facilitates comparison between languages by acknowledging an ordered range of patterns with each verb (see Fendel 2022a).

5 Summary and conclusion

Section 2 explained the three-partite distinction we make between a surface level, an underlying structural level and a conceptual level, the three-partite distinction we make between standard patterns, variations and deviations, and the tree-partite distinction we make between formulaic, semi-formulaic and free contexts. Section 2 furthermore drew attention to the fact that not every nonstandard pattern we find is due to language contact and not every language-contact phenomenon is an instance of bilingual interference.

 Section 3 looked at types of complements therein adopting Lehmann's (1988) continuum and drew attention to the fact that our morphosyntactic distinction between finite and infinite patterns in Greek and our semantic distinction between factive and prospective patterns do not map onto each other one-to-one. Section 3 furthermore introduced the methodology applied here when grouping tokens into classes by means of their argument and participant structures and when grouping complementation patterns into types, that is the distinction between collocations, idioms and formulae. The approach taken here is in essence similar to the idea of verb profiles advanced for modern languages such as French and German. Finally, Section 3 drew attention to three fundamental differences between Greek and Coptic, differences that affect the realisation of complementation patterns in these two languages. The three aspects are the function of word order, the use of rectional and combinatory elements with verbs and the morphosyntactic encoding of complementation patterns.

 Section 4 introduced the select corpus of texts, a corpus of private letters belonging to bilingual papyrus archives, applied the collocation-idiom-formula approach to γράφω, οἶδα and θαυμάζω and tabulated the distribution of variations and deviations for the select corpus of texts. Section 4 furthermore evaluated the usefulness of the approach.

 In conclusion, the chapter suggests acknowledging an ordered range of complementation patterns for Greek verb phrases (lexical verbs or verbal multi-word expressions) in the form of verb profiles. These verb profiles are multi-layered, in that they are built not only based on the c- and f-structures of a verb

phrase but also based on extralinguistic factors impacting on people's linguistic choices (i.e. variant and deviant patterns). Verb profiles can include patterns that are limited in their contextual applicability, as shown for γράφω, οἶδα and θαυμάζω, but this does not have to be the case. Regular internal processes affecting the number of participants of an event and their roles may also underlie verb profiles, as discussed for ἀξιόω, λέγω and θέλω.

References

Adams, James 2003. *Bilingualism and the Latin language*. Cambridge: Cambridge University Press.

Adams, Sean 2013. 'Atticism, classicism, and Luke-Acts: Discussions with Albert Wifstrand and Loveday Alexander', in Stanley E. Porter and Andrew W. Pitts (eds.), *The Language of the New Testament: Context, History, and Development.* Leiden – Boston: Brill, pp. 91–111.

Bagnall, Roger S. 2007. 'Reflections on the Greek of the Narmouthis Ostraka', in Mario Capasso and Paola Davoli (eds.), *New archaeological and papyrological researches on the Fayyum: Proceedings of the international meeting of Egyptology and papyrology, Lecce, June 8th-10th 2005.* Galatina: Congedo, pp. 15–21.

Bergeron, Marianne 2015. 'Pots and People: Greek Trade and Votive Rituals at Naukratis', in Damian Robinson and Frank Goddio (eds.), *Thonis Heracleion in Context: The Maritime Economy of the Egyptian Late Period, Proceedings of the Conference in the University of Oxford, 15-17 March 2013.* Oxford: Oxford University Press, pp. 267–81.

Bentein, Klaas 2017. 'Finite vs. non-finite complementation in Post-classical and Early Byzantine Greek: Towards a pragmatic restructuring of the complementation system?', *Journal of Greek Linguistics*, *17*(1): 3–36. https://doi.org/10.1163/15699846-01701002

Birney, Megan E., Rabinovich, Anna, Morton, Thomas A., Heath, Hannah, and Ashcroft, Sam 2020. 'When Speaking English Is Not Enough: The Consequences of Language-Based Stigma for Nonnative Speakers', *Journal of Language and Social Psychology*, *39*(1): 67–86.

Bortone, Pietro 2010. *Greek prepositions: From antiquity to the present.* Oxford: Oxford University Press.

Bresnan, Joan, Asudeh, Ash, Toivonen, Ida and Wechsler, Stephen 2015. *Lexical Functional Syntax* (2nd ed.). Chicester: John Wiley & Sons.

Bucking, Scott 2012. 'Towards an archaeology of bilingualism', in Patrick James and Alex Mullen (eds.), *Multilingualism in the Graeco-Roman Worlds.* Cambridge: Cambridge University Press, pp. 225–64. https://doi.org/10.1017/CBO9781139012775.011

Butler, Yuko G. and Hakuta, Kenji, 2004. 'Bilingualism and Second Language Acquisition', in Tej K. Bhatia and William C. Ritchie (eds.), *The Handbook of Bilingualism.* Malden, MA: Blackwell, pp. 114–44. https://doi.org/10.1002/9780470756997.ch5

Celano, Giuseppe 2013a. 'Argument-focus and predicate-focus structure in Ancient Greek', *Studies in Language*, *37*(2): 241–66.

Celano, Giuseppe 2013b. 'Word Order', in *Encyclopedia of Ancient Greek Language and Linguistics*. Leiden – Boston: Brill.

Cheng, Lauretta S.P., Burgess, Danielle, Vernooij, Natasha, Solís-Barroso, Cecilia, McDermott, Ashley, and Namboodiripad, Savithry 2021. 'The Problematic Concept of Native Speaker in Psycholinguistics: Replacing Vague and Harmful Terminology With Inclusive and Accurate Measures', *Frontiers in Psychology, 12*: 1–22.

Choat, Malcolm 2007. 'Epistolary formulae in early Coptic letters', in Nathalie Bosson and Anne Boud'hors (eds.), *Actes du huitième Congrès international d'études coptes* (Leuven: Peeters), 667–8.

Choat, Malcolm 2009. 'Language and Culture in Late Antique Egypt', in Jutta Raithel and Philip Rousseau (eds.) *A Companion to Late Antiquity*. Chichester: Wiley-Blackwell*,* pp. 342–56. https://doi.org/10.1002/9781444306101.ch23

Choat, Malcolm 2010. 'Early Coptic epistolography', in Arietta Papaconstantinou (ed.), *The Multilingual experience in Egypt, from the Ptolemies to the Abbasids*. Farnham: Ashgate, pp. 153–78.

Choat, Malcolm 2012. 'Coptic', in Christina Riggs (ed.), *The Oxford handbook of Roman Egypt.* Oxford: Oxford University Press, pp. 581–93.

Clackson, Sarah 2004. 'Papyrology and the utilization of Coptic sources', in Petra Sijpesteijn and Lennart Sundelin, *Papyrology And The History Of Early Islamic Egypt.* Boston – Leiden: Brill, pp. 21–44.

Clarysse, Willy 1993. 'Egyptian Scribes writing Greek', *Chronique d'Egypte*, *68* (135–136): 186–201. https://doi.org/10.1484/J.CDE.2.308932

Clarysse, Willy 2010. 'Linguistic diversity in the archive of the engineers Kleon and Theodoros', in Trevor V. Evans and Dirk D. Obbink (eds.), *The language of the papyri.* Oxford: Oxford University Press, pp. 35–50.

Crespo, Emilio 2007. 'The Linguistic Policy of the Ptolemaic Kingdom', in M.B. Chatzopoulos and Vassiliki Psilakakou (eds.), *Phōnēs Charaktēr Ethnikos: Actes Du Ve Congrès International de Dialectologie Grecque (Athènes 28-30 Septembre 2006)*, pp. 35–49.

Cristofaro, Sonia 2008. 'A constructionist approach to complementation: Evidence from Ancient Greek', *Linguistics 46*: 571–606. https://doi.org/10.1515/LING.2008.019

Crystal, David 2008. *A Dictionary of Linguistics and Phonetics* (6th edn). Malden, MA; Oxford: Blackwell.

Cromwell, Jennifer and Grossman, Eitan 2017. *Scribal Repertoires in Egypt from the New Kingdom to the Early Islamic Period*. Oxford: Oxford University Press.

Dahlgren, Sonja 2016. 'Towards a definition of an Egyptian Greek variety', *Papers in Historical Phonology*, *1*: 90–108. https://doi.org/10.2218/pihph.1.2016.1695

Dahlgren, Sonja 2017. *Outcome of long-term language contact: Transfer of Egyptian phonological features onto Greek in Graeco-Roman Egypt* (Thesis, University of Helsinki).

Davis, Boyd H., and Maclagan, Margaret 2020. 'Signposts, Guideposts, and Stalls: Pragmatic and Discourse Markers in Dementia Discourse', in Trini Stickle (ed.), *Learning from the Talk of Persons with Dementia: A Practical Guide to Interaction and Interactional Research*. Cham: Palgrave Macmillan, pp. 63–83.

Debaisieux, Jeanne-Marie 2004. 'Les Conjonctions de Subordination: Mots de Grammaire Ou Mots Du Discours ? Le Cas de Parce Que', *Revue de Sémantique et Pragmatique 15–16*: 51–67.

Depauw, Mark 1997. *A companion to demotic studies*. Bruxelles: Fondation égyptologique reine Élisabeth.

Depauw, Mark 2012. 'Language use, literacy and bilingualism', in Christina Riggs (ed.), *The Oxford Handbook of Roman Egypt*. Oxford: Oxford University Press, pp. 493–506.

Engsheden, Åke 2008. 'Differential object marking in Sahidic Coptic', in Folke Josephson and Ingmar Söhrman (eds.), *Interdependence of diachronic and synchronic analyses*. Amsterdam: John Benjamins, pp. 324–44.

Engsheden, Åke 2018. 'Verbal semantics and differential object marking in Lycopolitan Coptic', in Ilja A. Seržant and Alena Witzlack-Makarevich (eds.), *Diachrony of differential argument marking*. Berlin: Language Science Press, pp. 153–82.

Erman, B. 2001. 'Pragmatic Markers Revisited with a Focus on You Know in Adult and Adolescent Talk', *Journal of Pragmatics*, *33*(9): 1337–59.

Evans, Trevor 2012a. 'Complaints of the natives in a Greek dress. The Zenon archive and the problem of Egyptian interference', in Patrick James and Alex Mullen (eds.), *Multilingualism in the Graeco-Roman worlds*. Cambridge: Cambridge University Press, pp. 106–23.

Evans, Trevor 2012b. 'Linguistic and stylistic variation in the Zenon archive', in Martti Leiwo, Hilla Halla-Aho, and Marja Vierros (eds.), *Variation and chance in Greek and Latin*. Helsinki: Finnish Institute at Athens, pp. 25–42.

Fendel, Victoria 2021. 'The Missing Piece in the Jigsaw Puzzle: A Psycholinguistic Account of the Beginnings of the Coptic Alphabet', in Anna P. Judson and Amalia E. Gnanadesikan (eds.), *Diversity in Writing Systems Embracing Multiple Perspectives*. Amsterdam: John Benjamins, pp. 198–228.

Fendel, Victoria 2022a. 'Natural Language Use and Bilingual Interference: Verbal Complementation Patterns in Post-Classical Greek', in Michele Bianchoni and Marta Campano (eds.), *Ancient Indo-European languages between linguistics and philology: Contact, variation, and reconstruction*. Leiden – Boston: Brill, pp. 166–195.

Fendel, Vicotria 2022b. *Coptic Interference in the Syntax of Greek Letters from Egypt*. Oxford: Oxford University Press.

Fischer-Bovet, Christelle 2014. *Army and Society in Ptolemaic Egypt*. Cambridge: Cambridge University Press.

Fournet, Jean-Luc 2009. 'Esquisse d'une anatomie de la lettre antique tardive d'après les papyrus', in R. Delmaire, J. Desmulliez and P. Gatier (eds.), *Correspondances. Documents pour l'histoire de l'Antiquité tardive. Actes du colloque international, Université Charles-de-Gaulle-Lille 3, 20-22 novembre 2003*. Lyon: Maison de l'Orient et de la Méditerranée-Jean Pouilloux, pp. 23–66.

Fournet, Jean-Luc 2011. 'The Multilingual Environment of Late Antique Egypt: Greek, Latin, Coptic, and Persian Documentation', in Roger Bagnall (ed.), *The Oxford Handbook of Papyrology*. Oxford: Oxford University Press, pp. 418–51.

Fournet, Jean-Luc 2019. *The Rise of Coptic: Egyptian versus Greek in Late Antiquity*. Princeton – Oxford: Princeton University Press.

Gignac, Francis 1976. *A grammar of the Greek papyri of the Roman and Byzantine periods*. Milan: Istituto editoriale cisalpino-La Goliardica.

Gries, Stefan and Stefanowitsch, Anatol 2004. 'Extending collostructional analysis', *International Journal of Corpus Linguistics*, *9*(1): 97–129.

Gross, Maurice 1984. 'Lexicon-grammar and the Syntactic Analysis of French', *Proceedings of the 10th International Conference on Computational Linguistics*, pp. 275–82. https://doi.org/10.3115/980431.980549

Grossman, Eitan 2009. 'Periphrastic perfects in the Coptic dialects', *Lingua Aegyptia*, *17*: 81–118.

Grossman, Eitan 2015. 'No case before the verb in Coptic', in Eitan Grossman, Martin Haspelmath and Tonio Sebastian Richter (eds.), *Egyptian-Coptic Linguistics in Typological Perspective* (Berlin: Mouton De Gruyter), pp. 203–25.

Grossman, Eitan 2019. 'Language-Specific Transitivities in Contact: The Case of Coptic', *Journal of Language Contact*, *12*(1): 89–115. https://doi.org/10.1163/19552629-20180001

Grossman, Eitan, Dils, Peter, Richter, Tonio Sebastian and Schenkel, Wolfgang 2017. *Greek Influence on Egyptian-Coptic: Contact-Induced Change in an Ancient African Language (DDGLC Working Papers 1)*. Hamburg: Widmaier Verlag.

Hall, Jonathan M. 2014. *A History of the Archaic Greek World: Ca. 1200-479 BCE* (2nd edn). Malden, MA: Wiley-Blackwell.

Hanks, Patrick 1996. 'Contextual Dependency and Lexical Sets', *International Journal of Corpus Linguistics*, *1*(1): 75–98.

Hanks, Patrick 2013. *Lexical analysis: Norms and exploitations*. Cambridge, MA: MIT Press.

Hasznos, Andrea 2006. 'A Case Where Coptic is more Syndetic than Greek', *Acta Antiqua Academiae Scientiarum Hungaricae*, *46*(1–2): pp. 91–7. https://doi.org/10.1556/AAnt.46.2006.1-2.11

Hasznos, Andrea 2012. *Graeco-Coptica: Greek and Coptic clause patterns*. Wiesbaden: Harrassowitz.

Hasznos Andrea 2017. 'Syntactic Patterns Used after Verbs of Exhorting', in E. Grossman, Peter Dils, Tonio Sebastian Richter and Wolfgang Schenkel (eds.), *Greek Influence on Egyptian-Coptic: Contact-Induced Change in an Ancient African Language (DDGLC Working Papers 1)*, Hamburg: Widmaier, pp. 237–64.

Hulleberg Johansen, Stine 2021. 'A Contrastive Approach to the Types of Hedging Strategies Used in Norwegian and English Informal Spoken Conversations', *Contrastive Pragmatics*, *2*(1): pp. 81–105.

Hult, Karin 1990. *Syntactic variation in Greek of the 5th century A.D.* Göteborg: Acta Universitatis Gothoburgensis.

James, Patrick 2008. *Retention and retreat: Complementary participles and infinitives with verbs of perception and declaration in the Roman and Byzantine documentary papyri* (Thesis, University of Cambridge). https://doi.org/10.17863/CAM.31210

Jiménez López, Maria Dolores 2016. 'Funciones semánticas alternantes y diátesis verbal: el caso de παρασκευάζω', in Esperança Borrel Vidal and Pilar Gómez Cardó (eds.), *Omnia Mutantur: Canvi, Transformació y Pervivència En La Cultura Clàssica, En Les Seves Llengües i En El Seu Llegat*. Barcelona: Edicions de la Universitat de Barcelona, pp. 199–210.

Keenan, James 2007. 'Byzantine Egyptian villages', in Roger Bagnall (ed,), *Egypt in the Byzantine world, 300–700*. Cambridge: Cambridge University Press, pp. 226–43.

Kim, C. 2011. *'Grüße in Gott, dem Herrn' Studien zum Stil und zur Struktur der griechischen christlichen Privatbriefe aus Ägypten* (Thesis, University of Trier).

Kirk, A. 2012. *Word order and information structure in New Testament Greek* (Thesis, University of Leiden).

Kiss, Z. 2007. 'Alexandria in the fourth to seventh centuries', in Roger Bagnall (ed.), *Egypt in the Byzantine world, 300–700*. Cambridge: Cambridge University Press, pp. 187–206.

Koskenniemi, Heikki 1956. *Studien zur Idee und Phraseologie des griechischen Briefes bis 400 n. Chr.* Helsinki: sn.

Kraus, Thomas J. 2000. '(Il)Literacy in Non-Literary Papyri from Graeco-Roman Egypt: Further Aspects of the Educational Ideal in Ancient Literary Sources and Modern Times', *Mnemosyne* 53(3): 322–42.

Krause, Martin 1969. 'Die Testamente der Äbte des Phoibammon-Klosters in Theben', *Mitteilungen Des Deutschen Archäologischen Instituts. Abteilung Kairo*, *25*: pp. 57–67.

Labov, William 1991. *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.

Layton, Bentley 2011. *A Coptic grammar: With chrestomathy and glossary: Sahidic dialect* (3rd ed., rev.). Wiesbaden: Harrassowitz.

Lee, John 2013. 'The Atticist Grammarians', in Stanley E. Porter and Andrew Pitts (eds.), *Linguistic Biblical Studies*: *Vol. 6. The Language of the New Testament: Context, History, and Development*. Leiden – Boston: Brill, pp. 283–308.

Lehmann, Christian 1988. 'Towards a typology of clause linkage', in Sandra A. Thompson and John Haiman (eds,), *Clause combining in grammar and discourse* (Amsterdam: John Benjamins), pp. 181–225. https://doi.org/10.1075/tsl.18.09leh

Leiwo, Martti 2003. 'Scribes And Language Variation', in Leena Pietilä-Castren and Marjaana Vesterinen (eds.), *Grapta Poikila I*. Helsinki: Finnish Institute at Athens, pp. 1–11.

Luiselli, Raffaele 1999. *A study of high level Greek in the non-literary papyri from Roman and Byzantine Egypt* (Thesis, University of London).

Luraghi, Silvia 2003. *On the meaning of prepositions and cases: The expression of semantic roles in ancient Greek*. Amsterdam: John Benjamins.

MacCoull, Leslie 2009. *Coptic legal documents: Law as vernacular text and experience in late antique Egypt*. Tempe, Ariz.: [Turnhout, Belgium]: Arizona Center for Medieval and Renaissance Studies ; in collaboration with Brepols.

Matras, Yaron 2009. *Language contact*. Cambridge: Cambridge University Press.

Müller, Matthias 2012. 'Greek connectors in Coptic. A contrastive overview II', *Lingua Aegyptia*, *20*: 111–64.

Myers-Scotton, Carol 2002. *Contact Linguistics: Bilingual Encounters and Grammatical Outcomes*. Oxford University Press.

Myers-Scotton, Carol, and Jake Janice L. 2000. 'Four Types of Morpheme: Evidence from Aphasia, Code Switching, and Second-Language Acquisition', *Linguistics*, *38*(6): 1053–1100.

Namba, Kazuhiko 2010. 'Formulaicity in code-switching: Criteria identifying formulaic sequences', in D. Wood (ed.), *Perspectives on formulaic language: Acquisition and communication*. London: Continuum, pp. 129–50.

Penner, Zvi, and Bader, Thomas 1995. 'Issues in the Syntax of Subordination: A Comparative Study of the Complementizer System in Germanic, Romance and Semitic Languages with Special Reference to Bernese Swiss German', in Zvi Penner (ed.), *Topics in Swiss German Syntax*. Bern: Peter Lang, pp. 73–290.

Porter, Stanley E. and Pitts, Aandrew 2013. 'The Disclosure Formula in the Epistolary Papyri and in the New Testament: Development, Form, Function, and Syntax', in Stanley E. Porter and Andrew Pitts (eds.), *The Language of the New Testament: Context, History, and Development*. Leiden – Boston: Brill, pp. 421–38.

Quack, Joakim 2017. 'How the Coptic Script Came About', in Eitan Grossman, Peter Dils, Tonio Sebastian Richter, and Wolfgang Schenkel (eds.), *Greek Influence on Egyptian-Coptic: Contact-Induced Change in an Ancient African Language*. Hamburg: Widmaier, pp. 27–96.

Richter, Tonio Sebastian 2009. 'Greek, Coptic and the 'language of the Hijra': The rise and decline of the Coptic language in late antique and medieval Egypt', in Hannah M. Cotton (ed.), *From Hellenism to Islam: Cultural and Linguistic Change in the Roman Near East*. Cambridge: Cambridge University Press, pp. 401–46. https://doi.org/10.1017/CBO9780511641992.019

Richter, Tonio Sebastian 2014. 'Greek and Coptic in the Byzantine era', in James G. Keenan, J.G. Manning and Uri Yiftach-Firanko (eds.), *Law and legal practice in Egypt from Alexander to the Arab Conquest. A selection of papyrological sources in translation with introductions and commentary*. Cambridge: Cambridge University Press, pp. 134–44.

Robertson, Archibald T. 1919. *Grammar of the Greek New Testament in light of historical research* (3rd ed.). London: Hodder & Stoughton.

la Roi, Ezra 2022. 'Weaving Together the Diverse Threads of Category Change: Intersubjective Ἀμέλει "of Course" and Imperative Particles in Ancient Greek', *Diachronica*, *39*(2): 159–92.

Ruiz Yamuza, Emilia 2020. 'Insubordination in Ancient Greek? The Case of Ὥστε Sentences', in M. Leiwo, M. Vierros, and S. Dahlgren (eds.), *Papers on Ancient Greek Linguistics: Proceedings of the Ninth International Colloquium on Ancient Greek Linguistics (ICAGL 9) 30 August – 1 September 2018, Helsinki*. Helsinki: Societas Scientiarum Fennica, pp. 383–402

Rutherford, Ian 2010. 'Bilingualism in Roman Egypt? Exploring the archive of Phatres of Narmuthis', in Trevor V. Evans and Dirk D. Obbink (eds.), *The language of the papyri*. Oxford: Oxford University Press, pp. 198–207.

Schulz, Petra 2003. *Factivity: Its nature and acquisition*. Tübingen: Max Niemeyer.

Selinker, Larry 1972. 'Interlanguage', *International Review of Applied Linguistics*, *10*: 209–31.

Shisha-Halevy, Ariel 2007. *Topics in Coptic syntax: Structural studies in the Bohairic dialect*. Dudley, MA: Peeters.

Stolk, Joanne and Nachtergaele, Delphine 2016. 'Dative for Accusative Case Interchange in Epistolary Formulas in Greek Papyrus Letters', *Symbolae Osloenses*, *90*(1): 122–63.

Till, Walter C. 1964. *Die koptischen Rechtsurkunden aus Theben*. Wien: HBöhlaus Nachf, Kommissionsverlag der Österreichischen Akademie der Wissenschaften.

Tomasello, Michael 2003. *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Cambridge, Mass.: Harvard University Press.

Torallas Tovar, Sofia 2010. 'Greek in Egypt', in E. Bakker (ed.), *A Companion to the Ancient Greek Language*. Oxford: Oxford University Press, pp. 253–66.

van Minnen, Peter 2007. 'The other cities in Later Roman Egypt', in Roger Bagnall (ed.), *Egypt in the Byzantine world, 300–700*. Cambridge: Cambridge University Press, pp. 207–25.

Verstraete, Jean-Christophe 2007. *Rethinking the Coordinate-Subordinate Dichotomy: Interpersonal Grammar and the Analysis of Adverbial Clauses in English*. Berlin – New York: Mouton de Gruyter.

Vierros, Marja 2007. 'The language of Hermias, an Egyptian notary from Pathyris (c. 100 B.C.)', in Bernhard Palme (ed.), *Papyrologica Vindobonensia*: *Vol. 1. Akten des 23. Internationalen Papyrologenkongress, Wien, 22.-28. Juli 2001*, Vienna: Verlag der Österreichischen Akademie der Wissenschaften, pp. 719–23.

Vierros, Marja 2012. *Bilingual notaries in Hellenistic Egypt: A study of Greek as a second language*. Brussels: Comité Klassieke Studies, Subcomité Hellenisme, Koninklijke Vlaamse Academie van België voor Wetenschappen en Kunsten.

Villing, Alexandra 2015. 'Egyptian-Greek Exchange in the Late Period: The View Form Nokradj-Naukrati', in Damian Robinson and Franck Goddio (eds.),

Thonis Heracleion in Context: The Maritime Economy of the Egyptian Late Period, Proceedings of the Conference in the University of Oxford, 15-17 March 2013. Oxford: Oxford Centre for Martime Archaeology, pp. 229–46.

Weinert, Regina 2010. 'Formulaicity and usage-based language: Linguistic, psycholinguistic and acquisitional manifestations', in David Wood (ed.), *Perspectives on formulaic language: Acquisition and communication.* London: Continuum, pp. 1–20.

Wood, David 2010. *Perspectives on formulaic language: Acquisition and communication.* London: Continuum.

Wray, Alison 2009. 'Identifying formulaic language: Persistent challenges and new opportunities', in Roberta Corrigan, Edith A. Moravcsik, Hamid Oulali and Kathleen M. Wheatley (eds.), *Formulaic Language: Distribution and Historical Change* Vol. 1. Amsterdam: John Benjamins, pp. 27–52.

Wright, Brian J., and Ricchuiti, Tim 2011. 'From "God" (ΘΕΟΣ) to "God" (Ноүтє): A New Discussion and Proposal Regarding John 1:1c and the Sahidic Coptic Version of the New Testament', *The Journal of Theological Studies*, *62*(2): 494–512

Zakrzewska, Ewa 2006. 'The Hero, the Villain and the Mob', *Lingua Aegytia*, *14*: 325–46.

Zakrzewska, Ewa 2015. 'L⋆ as a Secret Language: Social Functions of Early Coptic', in Gawdat Gabra and Hany Takla (eds.), *Christianity and Monasticism in Middle Egypt.* Cairo: American University in Cairo Press, pp. 185–98.

Abbreviations

Acc	Accusative case (Greek)
Act	Active voice (Greek)
Aor	Aorist tense (Greek)
C	Collocation (Complementation pattern)
Dat	Dative case (Greek)
DirO	Direct Object
DOM	Direct Object Marking
F	Feminine gender
F	Formula (Complementation pattern)
Gen	Genitive case (Greek)
I	Idiom (Complementation pattern)
Ind	Indicative mood (Greek)
IndO	Indirect Object
Inf	Infinitive
Ladl	Laboratoire d'Automatique Documentaire et Linguistique
M	Masculine gender
Mid	Middle voice (Greek)
N	Neuter gender (Greek)
NP	Noun Phrase
Pas	Passive voice (Greek)
Perf	Perfect tense
Pl	Plural number
Pr	Present tense
Prt	Particle
Sbj	Subjunctive mood (Greek)
Sg	Singular number
Superl	Superlative (Greek)
VP	Verb Phrase

9. 'You Know Justice and Law and the Kind of Writing of the Notaries' (Rhet)or(ic)al Skills and Scribal Act in P.Col. inv. 600 (a.k.a. P.Budge), Coptic Transcript of a Hearing in front of an Arbitration Council

Tonio Sebastian Richter

1 P.Budge, a unique piece of documentary writing

P.Budge[1] conveys the words of a hearing that happened at Apollonopolis Anô / Edfu[2] in Upper Egypt in the course of arbitration proceedings between two conflicting parties in 646/7 CE, not long after the last battle fought over Alexandria between the Byzantines and the Arabs under Amr ibn al-ʿĀs. Matter in dispute is a house which Thecla, the deceased aunt of the deacon John, the suing party, had mortgaged more than two decades before to John's opponent, the peasant Philemon, for a loan of one solidus. In the original mortgage deed the transfer of ownership of this house 'after the (expiration of) the (repayment) date'

---

[1] P.Col. inv. 600 was edited, with the help of W.E. Crum, as 'The Budge Papyrus of the Columbia University' by A. Arthur Schiller (Schiller 1968) and reedited by Hasitzka 1993 (= KSB I 036). Both editions contain a number of shortcomings. A new edition and translation of P.Budge by Marzena Wojtczak and the present author is under preparation. Quotations from P.Budge in this Chapter are taken from this forthcoming edition.

[2] Edfu yielded the Greek papyri of the famous Papas archive (TM Arch. 170) from 'post-Byzantine' (to adopt Lajos Berkes's term) Egypt (ed. Rémondon, 1953, for its dating in the time of Muʿawiya see also Gascou and Worp 1982 and Foss 2009). Until recently it was not a prominent spot on the map of Coptic papyrology, except for two large papyri, P.Budge and the famous 'pepper monopoly' document (see below, 3.1 and fn. #16#). In 2009 the ostraca from the French-Polish Excavations of the 1930s were published (Bacot 2009, see also Delattre and Fournet 2011), most recently the edition of the substantial Coptic parts of the Papas archive have started (Boud'hors, Delattre et al. 2017); for a more detailed report on Coptic texts from Edfu, see Richter (2019: 160–2).

was stipulated, *metà tên protesmían* as the peasant puts it in a whole Greek phrase.[3] However, there seems to be an implicit, 'subliminal' consent between the disputing parties that one solidus was not an adequate purchase price (*iustum pretium*) for the house. Such feelings of inadequacy seem to underlie John's cause of action, and they were probably pivotal for Philemon's journey to Thekla's surviving children a year before the present hearing. There he paid another three solidi to Thekla's surviving descendants, the rightful heirs of the house, and received a deed of sale from them which would be crucial for the arbiters' final decision to dismiss John's case.[4] This sale document (P.BL inv. 2018 = SB VI 8987 from year 644/5 CE) as well as the deed of settlement (*dialysis*) in which John formally renounced his claim (P.BL inv. 2017 = SB VI 8988 from July 647 CE) formed part of Philemon's, the winning party's, family archive, where also P.Budge comes from.[5] If written protocolls were a regular outcome of hearings within out-of-court settlements, it was apparently not with any regularity that they ended up as parts of family archives; at any rate we hardly find this type of document in the papyrological record.[6] Why this happened to P.Budge we don't know. Whatever the reason was, it grants us the exceptional survival of a most remarkable piece of writing.

---

[3] The original acknowledgement of debt is called (lines 54, 55, 227) a ⲍⲩⲡⲟⲑⲏⲕⲏ ⲡⲣⲁⲥⲓⲥ *hypotʰêkê prasis*. It was most likely drawn up in Greek, not in Coptic: This is indicated by general considerations (see below, Section 2) as well as by the likely quotation from that document (line 74) ⲙⲉⲧⲁ ⲧⲏⲛ ⲡⲣⲟⲑⲉⲥⲙⲓⲁⲛ *metà tên protesmían*. An extant Coptic document of similar type, P.KRU 58, is called ⲁⲥⲫⲁⲗⲉⲥⲧⲉⲣⲟⲛ *aspʰalesteron* 'more secure (deed)', see Richter (2010: 53–4).

[4] According to C. 4,44,2 Impp. Diocletianus et Maximianus AA. Aurelio Lupo, a. 285, the *iustum pretium* should not fall below the half of the actual price of a purchase object ('…*si nec dimidia pars veri pretii*…'), as Cosima Möller (Berlin) explained me. Given Philemon's later payment of 3 solidi in addition to the one he had previously given as loan (which thus covered eventually ¼ of the price), there was clearly necessity to adjust the purchase price. Philemon himself refers to the concept of *iustum pretium* when he talks about the sale document he received from Thekla's heirs (line 276–7): ⲁⲛ⳨ ⲧⲁⲓⲕ[ⲁⲓⲁ] ⲧⲓⲙⲏ *a-n-tⁱ t-dik[aia] timê* 'we paid the *iustum pretium* (δίκαια τιμή)'. This is but one option to deal with the whole question; others will be discussed with by Jakub Urbanik (University of Warsaw) and Marzena Wojtczak (FU Berlin / University of Warsaw) in a juridical commentary on P.Budge which is under preparation.

[5] TM Arch. 190. A. Arthur Schiller, the first editor of P. Budge, was also the first one to recognise (Schiller 1964) the archival conection between P.Budge and two Greek documents *P.BL* inv. 2018 (*SB* VI 8987, deed of sale, 644/5 CE) and *P.BL* inv. 2017 (*SB* VI 8988, *dialysis* settlement, 647 CE) edited by Zilliacus (1940).

[6] See Kreuzsaler (2010: 21), who therefore argued that 'schiedsrichterliche Verfahren gewöhnlich nicht schriftlich aufgezeichnet wurden, sondern eine Protokollierung nur ausnahmsweise stattfand, und zwar vornehmlich bei Verhandlungen in koptischer Sprache'.

P.Budge is remarkable in several respects,[7] not least in terms of scribal practice and linguistics. In this Chapter I want to address the following points: I will start with general observations on the language choice in Philemon's archive within the context of changing habits in the contemporary judicial language (Section 2). I will then attempt to watch the scribe of P.Budge at work (Section 3). I will subsequently analyse the linguistic profile of the text and assess the significance of its features (Section 4), in order to eventually discuss the interaction between (rhet)or(ic)al and scribal acts in the making of P.Budge (Section 5).

2 Language choice in the archive of Philemon and the juridical language in seventh century Egypt

P.Budge belongs to the earlier private legal documents drawn up in the Coptic language[8] and it bears evidence for the growth of the Egyptian vernacular into this functional domain. The other documents from the same archival context, two notarial deeds (P.BL inv. 2017 and 2018) relating to the case of Thekla's house and a marriage contract of Philemon's daughter (P.BL inv. 2019), exhibit the traditional choice of Greek, the only language of juridical documents for more than three centuries after the ultimate decline of the Demotic instrument in Roman Egypt.[9] As Jean-Luc Fournet has shown, the shift of linguistic conventions in the juridical sphere after the mid-sixth century took its point of departure in the realm of arbitration, to move ahead from here to the realm of notarial deeds.[10]

While earlier Coptic legal documents are for the most part close renderings of contemporary Greek formularies,[11] the text of P.Budge gives us rare glimpses of an active, productive use of Coptic as a means to phrase judicial terms and concepts in the contemporary Egyptian vernacular language.

---

[7] By now, P.Budge has scarcely been studied as a source of the history of judicial thought and social history; it has mainly been dealt with by scholars working on legal practice, in particular that of out-of-couts settlements, such as Schiller (1971), Allam (1991) and (1992), Frantz Murphy (1993) and (2003), Kreuzsaler (2010) and Richter (2010).

[8] See Fournet (2010); MacCoull (2007) and (2013), Richter (2008: xxi-xxv, 16–25).

[9] On this decline see Depauw (2003), Lewis (1993), Muhs (2005), Richter (2013), Stadler (2011), Zauzich (1983).

[10] Fournet (2010) and (2019); on the earliest known Coptic notarial deed see Förster, Fournet and Richter (2012).

[11] See already Boulard (1913) and Steinwenter (1920).

A fascinating feature of P.Budge are the explicit reflections on the authenticity and legal force of written documents therein.[12] The issue of notarial authentification is important because John the deacon produced to the arbiters two letters, one of his deceased aunt and one of her son Mena, as alleged means of evidence, – evidence that is fiercely questioned by Philemon the peasant on grounds of diplomatic traits of written documents and how they matter for their legal validity. Although an illiterate man, as he states himself, he mentions several times criteria such as ϭⲓⲛⲥϩⲁⲓ *cin-shai* 'kind of writing' and ϭⲓⲛϣⲁϫⲉ *cin-šače* 'way of speaking', this is to say, diplomatic and textual features, to distinguish formal documents of legal validity (such as the deed of sale in his own possession) from informal and therefore legally invalid pieces of writing (such as the letters brought forward by John) (Ex. 1):

> **(1)** Lines 93–5 (Philemon):
> ⲁⲩⲱ ⲉϥⲧⲱⲟⲩⲛ ⲛ̄ϩⲉⲛⲕⲟⲟϩ` ⲛ̄ⲥ⟦ⲧ⟧ⲭⲓⲧⲁⲣⲓⲛ ⲉϥⲥⲡⲟⲩⲇⲁⲍⲉ ⲉⲣ̄ ϩⲉⲛϣⲁϫⲉ ⲉⲁⲩⲧⲃ̄ⲧⲱⲃⲟⲩ ⲙ̄ ⁹⁴| ⲙⲉ `·` ⲧⲉⲧⲛ̄ⲙ̄ⲡϣⲁ `ⲅⲁⲣ` ⲛ̄ⲉⲓⲙⲉ ⲉⲧϭⲓⲛϣⲁϫⲉ ⲙⲛ̄ ⲧϭⲓⲛⲥϩⲁⲓ̈ ⲛ̄ⲧϣⲟⲣⲡ̄ ⲛ̄ⲕⲟⲟϩ` ⲛ̄ⲥⲭⲓⲧⲁⲣⲓⲛ [ⲙ]ⲛ̄ [ⲧ]ϭⲓⲛϣⲁϫⲉ ⲛ̄ⲧⲙⲉϩ` ⲥⲛ̄ⲧⲉ ⲛ̄ ⁹⁵| ϣⲧⲁ`
> *auô e-f-tôoun n-hen-kooh n-s[t]kʰitarin e-f-spoudaze e-r-henšače e-a-u-tbtôbo-u m-* ⁹⁴| *me · tetn-mpša gar neime e-t-cin-šače mn t-**cin-shai** n-t-šorp n-kooh n-skʰitarin [m]n [t-]**cin-šače** n-t-meh snte n-* ⁹⁵| *šta*
> 'And he (John) exhibits fragments of drafts, making efforts to present some fabricated words as truth; but you areworthy to recognise the **kind of writing** of the first fragment and the **way of speaking** of the second snippet.'

No less than twenty times he ridicules John's letters by apostrophizing them as *sxedarion* 'drafts', or even worse, ⲕⲟⲟϩ` resp. ϣⲧⲁ ⲛ̄ⲥⲧⲓⲭⲁⲣⲓⲛ *kooh / šta n-sxidarion* 'fragments of drafts', and as 'having neither start nor end':[13]

---

[12] One of numerous *hapax* borrowings in P.Budge (see below, 4.3) is the phrase ⲇⲓⲁ ⲛⲟⲙⲓⲕⲟⲩ *dia nomikoú*, 'by notary, notarial'.

[13] These letters are *verbatim* quoted in P.Budge lines 24–8, 28–34 and 34–52 and are apparently not literally incomplete (although a *spatium* intentionally left by the scribe in line 27: ⲁϥϫⲟⲟⲥ ⲉⲣⲟⲛ (*spatium*) ⲙ̄ⲡⲉⲛϣⲁϫⲉ ⲛⲙ̄ⲙⲁϥ *afčoos eron (…) mpenšače nmmaf* 'He said to us: … but we didn't talk to him' may indeed indicate a gap or illegible passage in the original item), the attribute 'having neither start nor end' rather seems to refer to their lack of the initial and concluding *formulaic constituents of formal legal deeds* which by default would mark a writ as being issued by an accredited notarial authority.

(2) Lines 78–9 (Philemon):

ⲘⲀⲚⲈⲒ̈ⲔⲞⲞⳅ` Ⲛ̄ⲤⲦⲓⲬⲀⲣⲓⲚ^sic ⲚⲀⲒ̈ ⲈⲦⲈ ⲘⲚ̄- ⁷⁹| ⲦⲞⲨ ⲀⲣⲬⲎ` ⲞⲨⲆⲈ ⳅⲀⲎ` ⲣ̄ ⳽ⲀⲨ Ⲛ̄ⲦⲈⲣⲘⲞⲨⲀⲈ ⲚⲦⲀⲘⲓⲚⲈ ⲦⲰⲞⲨⲚ Ⲙ̄ⲘⲞⲞⲨ Ⲏ Ⲛ̄ⲤⲈⲦⲀⲀⲨ ⲚⲎⲦⲚ̄ ⳅⲰⲤ ⲆⲓⲔⲀⲓⲰⲘⲀ`

*ma-nei-**kooh n-stik^harin** nai ete mn-* ⁷⁹|*t-ou ark^hê oude haê r-šau nte-rmouae n-ta-mine tôoun mmo-ou ê n-se-taa-u nê-tn hôs dikaiôma*

'These fragments of drafts, having neither start nor end, are not worthy that (even) a peasant like me would exhibit them or hand them over to you, as if they were means of evidence.'

Time and again he reminds the arbiters:

(3) Lines 278–9 (Philemon):

ⲈⲦⲈⲦⲚ̄ⲤⲞⲞⲨ(Ⲛ) ²⁷⁹| Ⲙ̄ⲠⲆⲓⲔⲀⲓⲞⲚ ⲀⲨⲰ ⲠⲚⲞⲘⲞⲤ ⲘⲚ̄ Ⲧ6ⲒⲚ6ⳅⲀⲒ̈ Ⲛ̄ⲚⲚⲞⲘⲓⲔⲞⲤ Ⲛ̄ⲐⲎⲂⲀⲒ̈Ⲥ ⲘⲀⲖⲒⲤⲦⲀ ⲚⲀⲦⲀⲣⲅⲀⲦⲓⲀ

e-te-tn-soou(n) ²⁷⁹| *m-p-dikaion auô p-nomos mn t-cin-shai n-n-nomikos n-T^hêbais malista na-t-^Argatia*

'You know justice and law and the **kind of writing** of the notaries of the Thebais, especially those of Arkadia (*i.e., the administrative district of Middle Egypt*).'

The deacon John tries to play the same game by calling Philemon's deed of sale a 'void writ' and 'fabricated item' brought about 'for two *xestai* of oil'.[14] But this is not a bright idea since Philemon knows perfectly well, and repeatedly enumerates, the kinds and names of documents effecting the transfer of property, or at least the authorisation to act as someone's legal representative, whose possession John cannot boast about (Ex. 4):

(4) Lines 183–5 (Philemon):

ⲘⲚ̄ ⲖⲀⲀⲨ Ⲛ̄ⲆⲓⲔⲀⲓⲰⲘⲀ ⲈⲠⲰⲤ ⲠⲈ Ⲛ̄ⲦⲞⲞⲦⲞⲨ ⲈⲤⲈⲒⲣⲈ Ⲙ̄ⲘⲞⲞⲨ Ⲛ̄ⲬⲞⲈⲒⲤ ⲈⲠⲈⲚⲎⲒ̈ ⲞⲨⲆⲈ ⳅⲒⲦⲚ̄ **ⲠⲣⲀⲤⲓⲤ** ⲞⲨⲆⲈ ⳅⲒⲦⲚ̄ ¹⁸³| **ⲆⲰⲣⲈⲀ** ⲞⲨⲆⲈ ⳅⲒⲦⲚ̄ **ⲈⲔⲬⲰⲣⲎⲤⲓⲤ** ⲞⲨ^<ⲆⲈ> **ⲆⲓⲀⲐⲎⲔⲎ** ⲈⲀ⟦Ϥ⟧ⲨⳍⲰⲠⲈ ⲚⲞⲘⲒⲘⲰⲤ ⲞⲨⲆⲈ ⲞⲚ ⲠⲔⲈⲈⲚⲦⲞⲖⲒⲔⲞⲚ ⲈⲦⲈⳅⲀⲨⲤⲘⲚ̄Ⲧⳅ Ⲛ̄Ⲛ̄ⲣⲰ`ⲘⲈ´ ¹⁸⁴| ⲦⲀⲣⲞⲨⲣ̄ ⲠⲈⲨ⟦ⲣ̄⟧ⲠⲣⲞⲤⲰⲠⲞⲚ Ⲙ̄ⲠⲈⲦⲈⲦⲚ̄ⳅⲈ` ⲈⲞⲨⲞⲚ `ⲈⲠⲰⲤ ⲠⲈ´ Ⲛ̄ⲦⲞⲞⲦ⟦Ⲥ̄⟧ⲞⲨ `ⳅ´Ⲛ̄ ⲚⲀⲒ̈

---

[14] Line 219.224: ⲦⲀⲖⳅ Ⲛ̄ⲖⲀⲀⲨ ⲈⲦⲘ̄ⲘⲀⲨ…ⲦⲈⲒ̈ⲖⲀⲀⲨ Ⲛ̄ⲦⲞⲂⲦⲂ ⲀϤⲚ̄ⲦⲤ ⳅⲀ ⲤⲚⲀⲨ Ⲛ̄ⳝⲈⲤⲦⲎⲤ Ⲛ̄ⲚⲈⳅ *t-lek^s n-laau et-mma-u … tei-laau n-tobtb … a-f-nt-s ha-snau n-k^estês n-neh*. John seems to suspect the pay for notarial work to be a kind of bribery.

mn-laau n-dikaiôma e-pô-s pe ntoot-ou e-s-eire mmo-ou n-čoeis e-pe-n-êi oude
hitn **prasis** *oude hitn* **¹⁸³|dôrea** *oude hitn* **ekkᵇôrêsis** *ou⁽ᵈᵉ⁾ diatᵇêkê e-a-u-šôpe*
*nomimôs oude on p-ke-***entolikon** *ete-ša-u-smnt-s n-n-rôme* **¹⁸⁴|** *tar-ou-r pe-u-*
prosôpon mpe-tetn-he e-ouon e-pô-s pe ntoot-ou hn-nai

'No means of evidence pertaining to her (*Thecla*) is in their (*John's party's*)
hands, that she had made them lord over our house, neither by **deed of sale**
nor by **deed of donation** nor by **cession** or **will** drawn up legally, and not even
the **proxy document** that is usually issued for somebody to do someone else's
legal representation: You (*the arbiters*) haven't found one of those, pertaining to
her (*Thecla*), in their (*John' party's*) possession.'

He even knows precisely the triad of diplomatic features that grant legal force
to notarial documents – the issuer's stipulation, signatures of witnesses, and the
notary's completion note (Ex. 5):

(5) Lines 84–7 (Philemon):
ⲕⲁⲛ **⁸⁵|** ⲅⲁⲣ ⲁⲛⲅ̄ ⲟⲩⲣⲙ̄ⲟⲩⲁⲉ` †ⲥⲟⲟⲩⲛ ⲁⲛ ⲛ̄ϩⲱⲃ · ⲁⲗⲗ`ⲟⲩⲛ †ⲥⲱⲧⲙ̄ ϩⲓⲧⲛ̄
ⲛⲉⲧⲛⲟⲓ̈ ϫⲉ ⲙⲁⲭⲁⲣⲧⲏⲥ ⲉⲙⲛ̄ ϩⲩⲡⲟⲅⲣⲁⲫⲏ **⁸⁶|** ⲛ̄ϩⲏⲧϥ̄ ⲁⲩⲱ **ⲙⲁⲣⲧⲩⲣⲟⲥ** ⲁⲩⲱ
ⲕⲟⲙⲡⲗⲉⲩⲥⲓⲥ ⲛ̄**ⲛⲟⲙⲓⲕⲟⲥ** ⲣ̄ ϣⲁⲩ ⲛ̄ⲧⲉⲣⲱⲙⲉ ⲛ̄ⲧϥ̄ ⲛ̄ⲛⲁϩⲣⲛ̄ ⲇⲓⲕⲁⲥⲧⲏⲥ ⲛ̄ⲥⲉ **⁸⁷|**
ⲇⲓⲕⲁⲓⲟⲗⲟⲅⲉⲓ` ⲛ̄ϩⲏⲧϥ̄ · ⲙⲁⲗⲓⲥⲧⲁ ϣⲧⲁ` ⲛ̄ⲥⲧⲓⲭⲁⲣ[ⁱⁿˢⁱᶜ ⲛ̄]†ⲙ̄[ⲓ]ⲛ[ⲉ ⲉⲧⲉⲙ̄]ⲛ̄ⲧϥ̄
ⲁⲣⲭⲏ` ⲟ[ⲩ]ⲇⲉ̣ ϩⲁⲏ

kan **⁸⁵|** *gar ang ou-rmouae tⁱ-sooun an n-hôb all-oun tⁱ-sôtm hitn n-et-noi če ma-*
*kᵇartês e-mn-***hupograpᵇê** **⁸⁶|** *nhêt-f auô* **marturos** *auô* **kompleusis** *n-***nomikos**
r-šau nte-rôme nt-f nnahrn dikastês n-se- **⁸⁷|** *dikaiologei nhêt-f malista šta*
n-stikᵇar[inˢⁱᶜn-]tⁱ-m[i]n[e ete-m]nt-f arkᵇê o[u]de haê

'Even though I am only a peasant and have no understanding of the matter,
I heard after all by those who do have knowledge: A deed not exhibiting
signature, **witnesses** and **notarial kompleysis** is not worth to be produced
to the judge and to litigate trough it, let alone any snippets of drafts having
neither start nor end.'

These are the very features by which the arbiters will later assess and affirm the
authenticity of Philemon's deed of sale:

(6) Lines 200–3 (Arbiters):
ⲁⲩⲱ ⲛ̄ⲧⲉⲓ̈ϩⲉ ⲁⲩⲉⲡⲓⲇⲓⲇⲟⲩ ⲛⲁⲛ ⲛ̄ⲟⲩⲡ[ⲣⲁⲥⲓⲥ] **²⁰¹|** ⲉⲁⲩⲟⲩⲉⲓ`ˢⁱᶜ ϫⲉ ⲉⲗⲓⲥⲁⲃⲉⲧ`
ⲙⲛ̄ ⲕⲉⲟⲩⲁ` ϫⲉ ⲓ̈ⲉⲣⲁⲕⲓⲱⲛ ⲛ̄ϣⲏⲣⲉ ⲛ̄ⲑⲉⲕⲗⲁ` ⲙⲛ̄ ⲕⲉⲟⲩⲉⲓ` ϫⲉ ⲓ̈ⲟⲩⲥⲧⲓⲛⲁ ⲧⲥϩⲓⲙⲉ
ⲙ̄ⲙⲏⲛⲁ ⲡϣⲏⲣⲉ ⲛ̄ⲑⲉⲕ̣[ⲗⲁ] **²⁰²|** ⲥⲙⲛⲧⲥ ⲛⲁⲩ ϩⲙ̄ ⲡⲉⲙϫⲉ ⲧⲡⲟⲗⲓⲥ ⲇⲓⲁ ⲛⲟⲙⲓⲕⲟⲩ

ⲈⲨⲦ ⲈⲂⲞⲖ ⲚⲀⲨ ⲘⲠⲈⲒ̈ⲎⲒ̈ Ⲛ̄ⲦⲘⲒⲚⲈ … ⲦⲀⲒ̈
ⲞⲨⲚ ⲀⲚ� ⲌⲈⲢ ⲈⲢⲞⲤ · ⲈⲤⲬⲒ̀ ⲌⲨⲠⲞⲄⲢⲀⲪⲎ
ⲈⲤⲬⲒ̀ ⲘⲀⲢⲦⲨⲢⲞⲤ ⲈⲤⲬⲒ̀ ⲔⲞⲘⲠⲖⲈⲨⲤⲒⲤ

*auô n-tei-he a-u-epididou na-n n-ou-
p[rasis]* **201|** *e-a-u-ouei^(sic) če elisabet mn
ke-oua če ierakiôn n-šêre n-t^hekla mn
ke-ouei če ioustina t-shime m-mêna
p-šêre n-t^hek[la]* **202|** *smnt-s na-u hm-
pemče t-polis dia nomikou e-u-t^i ebol
na-u m-pei-êi n-t^i-mine … tai oun
a-n-he ero-s · e-s-či* **hupograp^hê** *e-s-či*
marturos *e-s-či* **kompleusis**

'And so they (Philemon's party) handed
over to us a deed of sale, notarially
issued to them by a certain Elisabeth
and a certain Hierakion, descendants
of Thekla, and a certain Justina, the
wife of the (late) Mena, the son of
Thekla, in the town of Oxyrhynchos,
according to which they sold them
the aforementioned house. … And
we found that it exhibits **signature**,
witnesses and **kompleusis**.'

3 Watching the scribe at work

3.1 The scribe of P.Budge

Although the name, title and precise
affiliation of the scribe of P.Budge remain
unknown to us, his work recommends him
as a very skilled clerk. His writing style is
bilinear, sloping, only slightly ligatured (see
Figure 1). His use of trema, invariably put
on consonantal *iota*, and his application
of supralinear strokes to mark syllabic
consonants and final vowels are remarkably
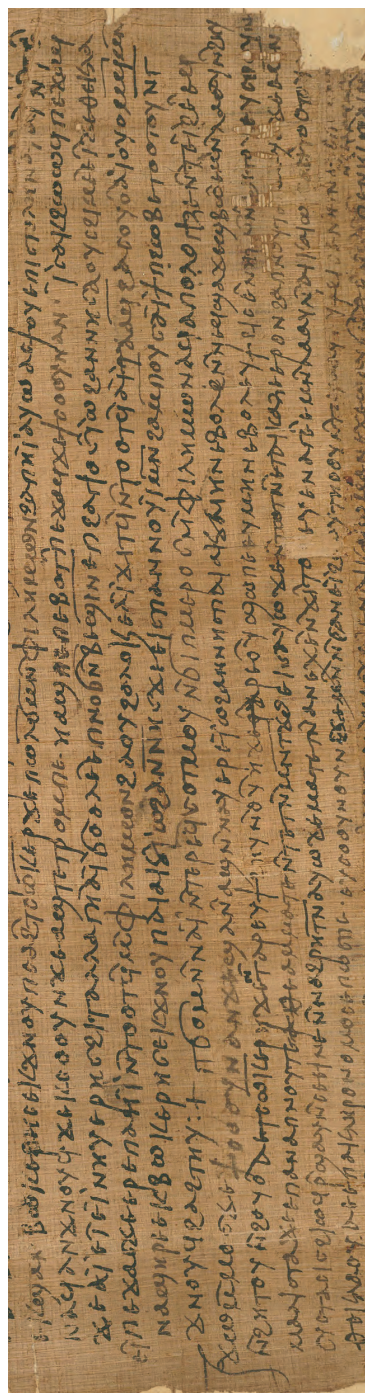
Figure 1.

consistent.[15] As Jennifer Cromwell has seen, yet another comprehensive Coptic document from Edfu is owed to his well-trained hand: BL Or. 8903, dated to indiction year 8 under the governor 'Abdallah ibn Sa'id = 649 CE, contains the acknowledgement of receipt of certain quantities of pepper by professional bodies (*koinôtai*) of crafts and trades of Apollonopolis magna;[16] it is addressed to Liberios, the pagarch. The two arbiters chairing the P.Budge hearing, Marinos and Markianos, were dignitaries (P.BL inv. 2017, 30–1: περιβλέπτοι πολιτευομένοι 'respectable *curiales*') of Apollonopolis Magna.[17] It seems, therefore, likely that the scribe of the two documents was a professional clerk affiliated to the municipal body of Apollonopolis magna.

3.2 The making of P.Budge

P.Budge is a complete papyrus scroll made up of eight papyrus leaves, each of about thirty centimeters height, together with a front paper stamped with false script and an end paper summing up to a length of 264 centimeters. Only thirty centimeters remained blank when the scribe was done with the text. To read aloud its 286 lines might have taken about forty-five minutes.

This large text was clearly written in one run: The several sections of the hearing – three speeches delivered by John, three responses given by Philemon, the transcript of the two letters produced by John, and the interrogation of three witnesses – are in no way separated in the layout of P.Budge: They follow one another, even within a line, in continuous script conducted with the same pen and the same ink, without traces of a new start; the *coronides* or *paragraphoi* put at the left margin to distinguish them may themselves be later additions (see Figure 1).

This observation is of significance since textual evidence tells us that the *hearing* was not done in one run but extended over several weeks or even months. In his second plea Philemon mentions (line 170) the period of four days that passed before John even admitted the invalidity of his letters. John says (line 207): 'For five months I have been troubling your lordships' And the three witnesses named by the suing party had apparently to be summoned during the hearing.

---

[15] This token of scribal skills is blurred by the misrepresentation of supralinear signs in both of the existing printed editions of P.Budge.

[16] BL Ms.Or. 8903 = SBKopt I 242 = TM 8798, ed. princ. by Crum (1925), reed. by Hasitzka in SBKopt. I 242. The governor Abdelas and the 8th indiction year allow to date this text exactly to 649 CE, see. J. Gascou and K.A. Worp, 'Problémes de documentation apollinopolite,' *ZPE* 49 (1982), 83–95.

[17] In P.Budge they call themselves 'we'; the two parties consistently refer to them by the honorific title ⲧⲉⲧⲛ̄ⲡⲉⲣⲓⲃⲗⲉ̂ⲭⲉ (ⲛ̄ϫⲟⲉⲓⲥ) *tetn-perible(ptotês) (n-čoeis)* 'you respectable (lords)'.

We may, therefore, take it for granted that P.Budge is *not the original transcript* of the parties' spoken utterances, but a later recension of the words originally produced and immediately transcribed during the days of the hearing. This conclusion ought to have implications for the issue whose language we are dealing with (see below, 4).[18]

On the other hand, P.Budge is apparently *the original version of this recension*, not a later copy of it. This is suggested by the final signatures, John's autograph signature (second hand, line 283–4) and Philemon's name written by an amanuensis (third hand, line 284–5), as well as by many corrections, changes and additions subsequently done by the scribe himself. These include the insertion of missing letters and words and a whole layer of revision running from line 1 up to line 179.[19] Also the addition of *coronides* at the left margin to index separate sections of the text might be part of the subsequent processing of the hearing protocol.

4 The linguistic profile of P.Budge

4.1 The point at issue

The very nature of the content of P.Budge, extensive utterances of (supposedly) oral speech rendered into written Coptic, suggests that traces of spoken language – in fact of several varieties of spoken language – would be conveyed by it, though not necessarily of a genuinely colloquial register but rather of a rhetorically-shaped oral performance.

In the light of what we said above (3.2), the precise point at issue is, therefore, the extent and kind of impact of the scribe's linguistic repertoire on the linguistic outcome after the underlying (presumably) oral speeches had been further processed into the written text of P.Budge. As we saw before, there must have been a first stage in the record of the hearing, previous to the version extant in P.Budge. One may *a priori* assume that the transformation of transcript notes taken during the hearing into the continuous text of P.Budge might, unconsciously or

---

[18] The different voices in P.Budge include John the deacon (ll. 5–23, 113–47, 204–32), Philemon the peasant (ll. 52–112, 162–200, 232–83), Thekla and her son Mena via their letters (ll. 24–52), three witnesses summoned by John's request (ll. 149–161) and the arbiters (ll. 1–5, 23f., 52f., 112f., 147–9, 161f., 200–4, 232).

[19] This revision aimed at a systematic change of pronouns referring to the accused party from singular (Philemon only) to plural (Philemon and his wife), thus from 'he' to 'they' if this party is referred to, and from 'I' to 'we' if the party is talking.

intentionally, have triggered replacements of spoken language features from the opponents' repertoires by written language features from the scribe's repertoire. But what does the linguistic profile of P.Budge tell us about this issue?

4.2 Morphophonological and syntactic features

The Coptic of P.Budge is an overall Sahidic Coptic idom, but displays a repertoire of grammatical features deviating from what we conceptualise as 'standard' or 'literary' Sahidic[20].

Some of them belong to a spectrum typical for South Egyptian Coptic, as otherwise attested in the southern Coptic dialects such as Akhmimic or in documentary texts from the Theban area, such as:

- The choice of 2nd Future (ϥⲛⲁⲥⲱⲧⲙ *e-f-na-sôtm*) rather than 3rd (ⲉϥⲉⲥⲱⲧⲙ *e-f-e-sôtm*) for injunctive future and generally for the extension of ϫⲉ(ⲕⲁⲥ) *čekas* 'so that'.[21]

> **(7)** Lines 56–7 (Philemon):
> ⲉⲥϣⲁⲛⲧⲙ̄ⲁⲡⲟⲗⲟⲅⲓⲍⲉ ⲛⲁⲛ ϣⲁ ⲧⲉⲡⲣⲟⲑⲉⲥⲙⲓⲁ ⲉⲧⲙ̄ⲙⲁⲩ ⲉϥⲛⲁϣⲱⲡⲉ ⲛⲁⲛ
> ϣⲁⲃⲟⲗ
> *e-s-šan-tm-apologize na-n ša te-prothesmia et-mmau e-f-na-šôpe na-n šabol*
> 'if she does not repay us up to this term, it (*i.e., the house)* shall be ultimately ours'

> **(8)** Line 111 (Philemon):
> ϫⲉⲕⲁⲥ ⲉⲣⲉⲡϫⲟⲉⲓⲥ ⲓ̄ⲥ̄ ⲡⲉⲭ̄ⲥ̄ ⲛⲁ2ⲁⲣⲉ2` ⲉⲣⲱⲧⲛ
> *čekas ere-p-čoeis I(êsou)s pe-X(risto)s na-hareh erô-tn*
> '… so that the Lord may protect you …'

- The morphology of Coptic Negated Aorist: pre-nominal ⲙⲁ- *ma-* (instead of ⲙⲉⲣⲉ- *mere-*), pre-pronominal ⲙⲁ⸗ *ma⸗* (instead of ⲙⲉ⸗ me⸗), *passim*.[22]
- The morphologically full ('absolute', unbound) form of the infinitive in

---

[20] For this standard see, e.g. Layton (2011), Shisha-Halevy (1988) and (1991).

[21] Shisha-Halevy (1976: 353–66); Richter (2008: 85–7); in Akhmimic: Till (1928, 146f., §127).

[22] Crum, *Coptic Dictionary* 182b; in Akhmimic: Till (1928, 145, §125); Nagel (1991, 23b).

bound construction with nominal object,[23] such as ⲉⲓⲛ- *ein-*,[24] ⲥⲱⲧⲙ- *sôtm-*,[25] ⲕⲱ- *kô-*,[26] ⲟⲩⲱⲣϩ- *ouôrh-*,[27] ⲡⲱⲗϭ- *pôlc-*,[28] ⲧⲱⲣⲡ- *tôrp-*.[29]

- The full (pre-pronominal) form of the dative preposition (ⲛⲁ- *na-*) before noun:[30]

(9) Line 256 (Philemon):
ⲡⲕⲉⲉⲛⲧⲟⲗⲓⲕⲟⲛ ⲉⲧⲉϣⲁⲣⲉ ⲛ̄ⲣⲱⲙⲉ ⲛⲓⲙ ⲥⲙⲛ̄ⲧ̄ϥ ⲛⲁ`ⲛ´ⲣⲱⲙⲉ̀
p-ke-entolikon ete-šare-n-rôme nim smnt-f na-n-rôme
'… also the mandatum that everyone would usually draw up for someone …'

- The morphology of the possessive article with 3rd plural possessor ⲡⲟⲩ- *p-ou-* / ⲛⲟⲩ- *n-ou-* 'their' besides ⲡⲉⲩ- *pe-u-* (once) and ⲛⲉⲩ- *ne-u-* passim):[31]

(10) Line 256 (Philemon):
ⲧⲁⲣⲟⲩⲣ̄ **ⲡⲟⲩⲡⲣⲟⲥⲱⲡⲟⲛ**
*tarou-r **p-ou**-prosôpon*
'…that they do **their** proxy…'

as opposed to

line 17 (John):
ⲛ̄ⲧⲉⲧⲁⲅⲟ` **ⲡⲟⲩϣⲁϫⲉ**
*n-te-tauo **p-ou**-šače*
'…that you say **your** (*fem.*) word…'

---

[23] Crum (1926, 250).

[24] Line 99: ⲁⲩⲉⲓⲛ` ⲛⲉⲓⲥⲭⲓⲧⲁⲣⲓⲛ *a-u-ein-nei-skʰitarin* 'they brought these drafts'; line 107: ⲁⲩⲉⲓⲛ` ⲛⲉⲓⲕⲟⲟϩ` *a-u-ein-nei-kooh* 'they brought these fragments'; line 148: ⲡⲇⲓⲁⲕⲱⲛ ⲉⲓⲛ ⲡⲁⲛⲛⲟⲩⲓ̈ *p-diakôn ein-Pannoui* 'the deacon brought Pannui'.

[25] Line 23: ⲧⲉⲧⲛ̄ⲥⲱⲧⲙ̄ ⲡⲉⲛϩⲱⲃ *tetn-sôtm-pen-hôb* 'you heard our case.'

[26] Line 7: ⲉⲁⲩⲕⲱ ⲧⲕⲟⲩⲓ̈ ⲛ̄ϣⲉⲉⲣⲉ *e-a-u-kô-t-koui n-šeere*, line 47: ⲉϣⲱⲡⲉ ⲕⲛⲁⲕⲱ` ⲡⲛⲟⲃⲉ *ešôpe k-na-kô-p-nobe* 'if you will leave the sin'.

[27] Line 10: ⲛ̄ⲧⲉⲣⲟⲩⲟⲩⲱⲣϩ̄ ⲧⲡⲟⲗⲓⲥ *nter-ou-ouôrh-t-polis* 'when they left the town'; the verb ⲟⲩⲱⲣϩ *ouôrh* itself is an Upper Egyptian lexical isogloss: Crum 1926, 253f.

[28] Line 41: ϣⲁⲛⲧⲁⲡⲱⲗ̄ϭ ⲡⲁϩⲱⲃ *šant-a-pôlc-pa-hôb*.

[29] Line 250: ⲉⲧⲱⲣⲡ̄ ⲡⲉⲧⲉⲡⲱⲛ ⲡ[ⲉ] *e-tôrp-p-ete-pô-n p[e]*.

[30] This is, according to Crum (1926, 248): 'perhaps the most exclusively Theban feature'.

[31] As in Akhmimic and the so-called Subakhmimic (Lycopolitan) dialects.

(11) Line 140 (John):

ⲚⲞⲨⲆⲒⲀⲐⲎⲔⲎ` … ⲚⲞⲨⲆⲰⲢⲈⲀ

*n-ou-diat^ʰêkê … **n-ou**-dôrea*

'…**their** wills ... **their** donations…'.

In contrast to those unmistakably southern isoglosses, two such features distinctively indicate Middle Egyptian Coptic:

    - A form ⲈⲦⲀ-Ϥ-ⲤⲰⲦⲘ *et-a-f-sôtm* (here in a non-standard form of second person sg.f. ⲈⲦⲀⲤⲰⲦⲘ *et-a-ø-sôtm*) of Second Perfect:

(12) Line 37 (Thekla's letter):

ⲈⲦⲀⲤⲘ̄Ⲛ ⲠⲢⲀⲤⲒⲤ ⲈⲠⲎⲒ̈ ⲚⲀⲒ̈ Ⲛ̄ⲦⲀ† ⲦⲀⲤⲪⲀⲖⲈⲒⲀ` ⲚⲈ

***eta**-smn prasis e-p-êi na-i nt-a-ti t-asp^ʰaleia ne*

'**once** you draw up a deed of sale for this house to me, I will give you back the debt instrument'

- Feminine gender assigned to the noun ϨⲞⲖⲞⲔ(ⲞⲦⲒⲚⲞⲤ) *holokotinos* 'solidus' (*a denomination of the Byzantine gold currency*):[32]

(13) Line 29 (Mena's letter):

ⲂⲰⲔ ⲈⲠϨⲀⲠ` Ⲛ̄ⲘⲘⲀϤ ⲈⲦⲂⲈ ⲦϨⲞⲖⲞⲔ̱ ⲈⲦⲈⲢⲞⲒ̈

*bôk e-p-hap nmma-f etbe **t**-holok(otinos) et-ero-i*

'go to court with him because of **the** (*fem.*) solidus'

(14) Line 31 (Mena's letter):

ⲈϢⲰⲠⲈ Ⲙ̄ⲠⲈⲦⲈϤϨⲞⲖⲞⲔ[]̣ [32]| ⲘⲞⲨϨ`

*ešôpe mpe-**te**-f-holok(otinos) mouh*

'if **his** (*fem.*) Solidus was not full…'

as opposed to four masculine instances of ϨⲞⲖⲞⲔ(ⲞⲦⲒⲚⲞⲤ) *holok(otinos)* in lines 56 (Philemon), 73 (Philemon), 159 (a witness), 223 (John).

Other peculiarities include:

- Attestations of protatic and apodotic ⲈϤⲤⲰⲦⲘ *e-f-sôtm*:[33]

---

[32] Feminine gender assignment to this word is limited to Coptic documents from Middle Egypt, see Crum, *Coptic Dictionary*, p. 140b s.v. ⲖⲞⲔⲞⳓ; Richter (2008, 332–3, ⲋ258).

[33] Crum (1926, 251); Shisha-Halevy (1973) and (1974); Grossman (2009).

(15) Lines 150, 156 and 160 (witnesses):

ⲉⲕϢⲁⲛ-/ⲉⲕⲂⲰⲔ ⲉⲢⲎⲤ ⲉⲕⲬⲚⲞⲨ

*e-k-(šan)-bôk e-rês **e-k**-čnou*

'If you go southwards you shall ask.'

- The occurrence of the etymologically improper first person of Causative Conjunctive:

(16) Line 19 (John):

ⲦⲀⲢⲓϨⲉ` ⲉⲢⲞⲞⲨ

tar-i-he ero-ou

'so that I shall find them.'

- The use of Aorist (ϢⲀϤⲤⲰⲦⲘ *ša-f-sôtm*, neg. ⲘⲀϤⲤⲰⲦⲘ *ma-f-sôtm*) as future tense:[34]

(17) Line 38 (Thekla)

ⲘⲀⲓⲦⲀⲀϤ ⲚⲀⲔ ⲈⲚⲈϨ · ⲀⲗⲗⲀ ϢⲀⲓⲦⲀⲀϤ ⲘⲠⲀⲤⲞⲚ ⲚϤϤⲒ ⲠⲢⲞⲞⲨϢ ⲚⲦⲀⲠⲢⲞⲤⲪⲞⲢⲀ

***ma**-i-taa-f na-k eneh alla **ša**-i-taa-f m-pa-son n-f-fi roouš n-ta-prospʰora*

'I will not give it to you ever, I will rather give it to my brother that he shall bring my (funerary) offerings.'.

- The reduplication of the conjugation base (neg. Aorist ⲘⲀ-...ⲘⲀ⸗ *ma-...ma⸗*):[35]

(18) Line 178 (John):

ⲘⲀⲚⲈⲨϢⲀⲬⲈ ⲚⲀⲤⲨⲤⲦⲀⲦⲞⲚ ⲘⲀⲨϢⲪⲈⲖⲈⲒ` ⲘⲘⲞⲞⲨ

***ma**-ne-u-šače n-asustaton **ma**-u-ôpʰelei mmo-ou*

'These incoherent words will not benefit them.'

- 'Incongruent' subject pronoun (ⲦⲈ *te* instead of ⲠⲈ *pe*) in nominal sentence:[36]

---

[34] Future semantics is rooted in this pattern's meaning of general, common, habitual, timeless truth or action (Steindorff [1951: 160]) and tends to be(come) more prevalent in non-literary registers and later Coptic language.

[35] Bosson (2006) and Layton (2011, §332) ('Extraposition of the subject with its own extraposed base: Ⲁ-ⲠⲢⲰⲘⲈ ⲀϤ-ⲤⲰⲦⲘ').

[36] See Boud'hors & Shisha-Halevy (2012). As an alternative analysis, ⲦⲈ *te* may instead be the subject pronoun of an (unmarked) relative clause: ⲐⲈⲔⲗⲀ - ⲦⲤⲰⲚⲈ ⲘⲠⲈϤⲈⲒⲰⲦ ⲦⲈ - *tʰekla t-sône m-pe-f-eiôt te* 'Thecla who (lit. she) is the sister of his father'. In this case the syntactic anomaly

(19) Line 3 (arbiters):

ⲡⲁ ⲟⲩⲥϩⲓⲙⲉ ϫⲉ ⲑⲉⲕⲗⲁ ⲧⲥⲱⲛⲉ ⲙ̄ⲡⲉϥⲉⲓⲱⲧ ⲧⲉ

*pa ou-shime če tʰekla t-sône m-pe-f-eiôt **te***

'It [*i.e. the house*] is that of a woman, a certain Thekla, the sister of his [*i.e. Johns*] father.'

- 'Double' determination by definite article as well as (quantifier) ⲛⲓⲙ *nim*:[37]

(20) Line 256 (Philemon):

ⲡⲕⲉⲉⲛⲧⲟⲗⲓⲕⲟⲛ ⲉⲧⲉϣⲁⲣⲉ ⲛ̄ⲣⲱⲙⲉ ⲛⲓⲙ ⲥⲙⲛ̄ⲧϥ

*p-ke-entolikon ete-šare-**n**-rôme **nim** smnt-f na-n-rôme*

'… also the mandatum that everyone usually draws up …'

- Predicative noun of an interlocutive nominal sentence without article:[38]

(21) Line 81 (Philemon):

ⲁⲛⲅ̄ ⲣ̄ⲙⲟⲩⲁⲉˋ ⲛ̄ⲧⲟϥ ⲇⲉ ⲟⲩⲡⲟⲗⲓⲧⲏⲥ ⲡⲉ

ang-rmouae ntof de ou-politês pe

'I am (a) peasant, he however is a townsman'

as opposed to line 85 (Philemon):

---

would not be an incongruent subject pronoun but the lack of any subject pronoun of the framing nominal sentence ⲡⲁ ⲟⲩⲥϩⲓⲙⲉ … (ⲡⲉ) *pa oushime … (pe)* 'It (*the house*) (is) that of a woman'.

[37] This construction is fairly well attested in Demotic, see Janet H. Johnson, 'The Use of the Articles and the Generic in Demotic', in: S. Vleeming (Ed.), *Aspects of Demotic Lexicography*, Studia Demotica 1, Leuven 1987, 52; Robert S. Simpson, *Demotic Grammar in the Ptolemaic Sacerdotal Decrees*, Oxford 1996, 48; on Sahidic Coptic see Ariel Shisha-Halevy, *Coptic Grammatical Categories*, 144, n. 16: 'I know of two Sah. instances of ⲛⲓⲙ in compatibility with non-zero determiners: ⲟⲩϩⲟⲟⲩⲧ ⲛⲓⲙ (Ex 12.48 Bodmer) and ⲡⲃⲁⲧⲟⲥ ⲛⲓⲙ (NHC VII 11.10)'. In Bohairic the situation is different, see the examples in Crum, *CD* 225b: ⲡⲓⲣⲉⲙⲛⲧ̇ⲙⲓ ⲛⲓⲃⲉⲛ *piremnṫmi niben* (Nu 15,13, *Sah.*: ⲟⲩⲟⲛ ⲛⲓⲙ *ouon nim*); ⲡⲓⲱⲧ ⲛⲓⲃⲉⲛ *piôt niben* (Lev 3,3) ⲟⲩⲣⲱⲙⲓ ⲛⲓⲃⲉⲛ *ourômi niben* (Lev 21,18), ⲫⲏ ⲛⲓⲃⲉⲛ ⲉⲧⲛⲁ-… *pʰê niben etna-…* (Lev 6,11) etc.

[38] This construction, the standard pattern of interlocutive nominal clauses in Demotic, and still attested in the Old Coptic Papyrus Schmidt, line 8: ⲁⲛⲉⲕ ⲁϭⲣⲏⲛ *anek acrên* 'I am barren', is no longer found in Coptic where determination has become a constituent of the nexus, see Hans-Jakob Polotsky, *Grundlagen des Koptischen Satzbaus* I, § 27, and Wolf-Peter Funk, 'Formen und Funktionen des interlokutiven Nominalsatzes im Koptischen', *LOAPL 3* (1991), 33 f.

ⲁⲛⲅ̄ ⲟⲩ-ⲣ̄ⲙⲟⲩⲁⲉ`

*ang-**ou**-rmouae*

'I am **a** peasant.'

- Reduced cleft sentence with focused interrogative pronomoun directly followed by relative clause without ⲡ(ⲉ) *p(e)*, ⲧ(ⲉ) *t(e)*, ⲛ(ⲉ) *n(e)*:[39]

(22) Line 152 (arbiters):

ⲟⲩ` ⲉⲧⲥⲏ2`

ou et-sêh

'What is it that is written?'

- Preposition ⲛⲧⲟⲟⲧ⸗ *ntoot⸗* in pre-pronominal state with suffix pronoun not grammatically agreeing with the following *n*-linked object:

(23) Line 22 (John):

ⲛ̄ⲧⲟⲟⲧⲥ̄ ⲛ̄ⲛⲓⲙ

ntoot-s (3rd sg.f.) *n-nim* (3rd person indef. c.)

'… from whom?'

(24) Line 266 (Philemon):

ⲛ̄ⲧⲟⲟⲧⲥ̄ ⲛ̄ⲛⲁⲓ̈

ntoot-s (3rd sg.f.) *n-nai* (3rd pl.c.)

'… from those …'

(25) Lines 256–7 and 272 (both Philemon):

ⲛ̄ⲧⲟⲟⲧϥ̄ ⲛ̄ⲛⲁⲓ̈

ntoot-f (3rd sg.m.) *n-nai* (3rd pl.c.)

'… from those …'

- Relative nominal clauses with possessive pronoun (ⲡⲱ⸗ⲓ *pô-I*, ⲡⲱ⸗ⲕ *pô-k*, 'mine, yours etc.') as predicate lacking the subject pronoun ⲡⲉ *pe*:[40]

---

[39] Polotsky (1962, 424f). This reduced type of cleft sentence is fairly common in Bohairic. In Sahidic literary texts it is a rare variant with first and second-person independent pronouns in the first position.

[40] Kahle (1954, §189 [¶ 157A]). Possession-marking relative clauses without subject clitic, i.e., having lost clausal properties, can already be found in literary standards of early Coptic dialects such as Early Bohairic (B$_4$) and Manichaean Lycopolitan (L$_4$) and were getting lexicalised in

(26) Lines 235, 245, 259 (all Philemon):

ⲡⲉⲧⲉⲡⲱⲛ

p-ete-pô-n

'… what belongs to us …'

as opposed to line 244 (Philemon):

ⲉⲩϣⲓⲛⲉ ⲛ̄ⲥⲁ ϩⲁⲣⲡⲁⲍⲉ ⲙ̄ⲡⲉⲧⲉⲡⲱⲛ ⲡⲉ

*e-u-šine nsa harpaze m-p-ete-pô-n **pe***

'they seek to steal what belongs to us'

and line 250 (Philemon):

ⲉⲧⲱⲣⲡ̄ ⲡⲉⲧⲉⲡⲱⲛ ⲡ[ⲉ]

*e-tôrp p-ete-pô-n **p[e]***

'to steal what belongs to us'

(27) Line 68 (Philemon):

ⲙ̄ⲡⲉⲥⲕⲱ` ⲗⲁⲁⲩ ⲉⲡⲱⲥ ⲛ̄ϩⲏⲧⲟⲩ

mpe-s-kô laau e-pô-s nhêt-ou

'She did not leave anything of her belongings in them [*i.e., the rooms of the house*].'

as opposed to line 102 (Philemon):

… ⲉⲡⲱⲛ ⲡⲉ

*e-pô-n **pe***

'… belonging to us …'

and lines 166, 182, 196, 247, 254 (all Philemon):

… ⲉⲡⲱⲥ ⲡⲉ

*e-pô-s **pe***

'… beloning to her …'

---

later Bohairic into an innovative focalised possessor construction 'breaking into the noun phrase'
(Grossman unpubl.).

- Pre-pronominal form ϩⲓⲱ⸗ *hiô⸗* instead of ⲛ̄ϩⲏⲧ⸗ *nhêt⸗*:[41]

    (28) Line 61 (Philemon):
    ⲛ̄ⲧⲁⲥⲉⲡⲓⲧⲣⲉⲡⲉˋ ⲛⲁⲩ ϩⲓⲱⲟⲩ
    *nta-s-epitrepe na-u **hiô-ou***
    '… **through** which she had authorised them (*sc. John and his party*) …'

as opposed to line 247 (Philemon):

    ⲉⲥⲉⲡⲓⲧⲣⲉⲡⲉˋ ⲛⲁⲩ ⲛ̄ϩⲏⲧⲟⲩ
    *e-s-epitrepe na-u **nhêt-ou***
    '…that she had authorised them (*sc. John and his party*) **through** them (*sc. her letters*).'

    (29) Line 164–5 (Philemon):
    ⲉⲙ̄ⲛ̄ⲥⲩⲥⲧⲁⲥⲓⲥ ϩⲓⲱϥ
    *e-mn-sustasis **hiô-f***
    '… while no legal force is **in** them.'

as opposed to line 181 (Philemon):

    ⲉⲙ̄ⲛ̄ ⲗⲁⲁⲩ ⲛ̄ⲥⲩⲥⲧⲁⲥⲓⲥ ⲛ̄ϩⲏⲧⲟⲩ
    *e-mn-laau n-sustasis **nhêt-ou***
    '… while no legal force is **in** them.'

---

[41] This phenomenon has rarely been mentioned (but see Berkes and Vanthiegem [2019, 15], and Richter [2020, 156, fn. 16]) and has never been dealt with properly although it is not infrequent in later (non-standard) Coptic. While the *semantic closeness* of the two locative prepositions ϩⲛ- *hn-* 'in' and ϩⲓ- *hi-* 'on' allowed for some degree of *quid pro quo* in *ad hoc* expressions (see examples in Crum, *CD* 644a) as well as in phraseology (cf. ϩⲛ̄ⲧⲙⲏⲧⲉ *hn-t-mête* vs. ϩⲓⲧⲙⲏⲧⲉ *hi-t-mête* 'in midst'), the phenomenon under discussion — the ongoing merger of the prepositional paradigms ϩⲓ-/ϩⲓⲱⲱ⸗ *hi-/hiôô⸗* and ϩⲛ̄-/ⲛ̄ϩⲏⲧ⸗ *hn-/nhêt⸗* at the expense of the pre-pronominal form of the latter (thus ϩⲓⲱⲱ⸗ *hiôô⸗* partially or totally replacing ⲛ̄ϩⲏⲧ⸗ *nhêt⸗*) — becomes more salient in texts such as *P.Ryl.Copt.* 106, SBKopt. I 001 or BL Or 3669(1) where ϩⲓ-/ϩⲓⲱⲱ⸗ *hi-/hiôô⸗* numerically equals, or even prevails over ϩⲛ̄-/ⲛ̄ϩⲏⲧ⸗ *hn-/nhêt⸗*. The phonetic conflation of the pre-nominal forms ϩⲓ- *hi-* and ϩⲛ- *hn-* in something like /hĭ/ ≈ /hĕ/ might have triggered or at least reinforced an increasing merger.

(30) Line 173 (Philemon):

ⲉⲩϣⲁϫⲉ ⲁⲛ ϩⲓⲱⲟⲩ

*e-u-šače an **hiô**-ou*

'… that they don't speak **in** them.'

versus line 97 (Philemon):

ⲉⲩϣⲁϫⲉ ⲛ̄ϩⲏⲧⲟⲩ

*e-u-šače **nhêt**-ou*

'… while they speak **in** them.'

It should be stated that none of these traits, though some do have a 'colloquial flavour', as it were, are entirely alien to any kind of written Coptic so as to be taken straightly and strictly as features of 'spoken' language.

4.3. Lexical features — the usage of loaned function words in P.Budge

If the linguistic profile of P.Budge goes fairly well along with morphosyntactic isoglosses otherwise known from Upper Egypt, and generally from nonstandard Coptic, it is outstanding when it comes to its lexicon, in particular the Greek loanwords therein. The 286 lines of P.Budge contain altogether 961 token instances of 224 lexeme types of Greek origin. Compared to other Coptic texts, these numbers are not in themselves conspicuous. Striking is, however, the share of different parts of speech therein. Compared to a big source of Greek loanword usage in Coptic such as the *DDGLC* database,[42] the type proportion of adjectives, verbs, adverbs, conjunctions, prepositions and particles is significantly higher than the Coptic average. This discrepancy is especially large in the categories of functional language (conjunctions, prepositions and particles) which sum up to 13.8 per cent for P.Budge, as opposed to 5.4 per cent average in Coptic (see Table 1). If proof for the pronounced rhethorical style of John's and Philemon's speeches were needed, this observation may suffice.

---

[42] https://www.geschkult.fu-berlin.de/en/e/ddglc

Table 1. The linguistic profile of P.Budge: Type frequency and shares of parts of speech (DDGLC data 2022)

Part of Speech	P.Budge	Share	DDGLC	Share
Substantives	96	42,8%	2.385	60,5%
Adjektives	26	11,6 %	381	9,7%
Verbs	59	26,3%	738	18,7%
Adverbs	12	5,3 %	114	2,9%
Autosemantic items total	193	86,2%	3618	94,6%
Conjunctions	11	4,9%	92	2,3%
Prepositions	10	4,5%	40	1,0%
Particles	6	2,7%	76	1,7%
Relative pronouns	3	1,4%		
Demonstrative pronouns	1	0,4%		
Functional items total	31	13,8%	208	5,4%
Types total	224	100%	3826	100%

Another striking observation about the loaned vocabulary of P. Budge is that no less than forty-five items, the fifth part of it, are *hapax* or near-to-*hapax legomena* at least within the vast corpus of Coptic documentary texts. What does this observation mean?

A good deal of these *hapax* words are clearly motivated by the juridical and thematic particularities of P.Budge, namely terms such as ἀκαίρως 'untimely(?)', ἀκυρόω 'to invalidate', ἀντίδικος 'opponent (in a lawsuit)', ἀποδείκνυμι 'to give evidence', γενεαλογία 'descent', διὰ νομικοῦ 'by notary, notarial', διάκρισις 'decision', διατίθημι 'to inherit', ἐκχώρησις 'cession', ἐντολικόν 'proxy document', μετουσία 'participation', ὄψις '(legal) state, condition', προγενέστερος 'first-born', συγκρίνω 'to pass sentence', υἱοθεσία 'adoption', ὑποθήκη 'mortgage'.

Others, however, are not related to the topic of the text and cannot be motivated by any kind of semantic pertinence or even need. This is most obviously the case of function words, the kind of loaned lexemes whose types are statistically overrepresented in P.Budge. A closer look at some of those might be at place.

The particle **γε** is attested once in P.Budge as part of a Greek multi-word expression, **ⲉⲓ ⲅⲉ** *ei ge*. It is used by Philemon to rhetorically highlight his argument when he talks about …

(31) Line 263 (Philemon):

ⲚⲢⲰⲘⲈˋ ˋⲚⲦⲀⲨⲔⲀⲚⲰⲚⲓⲌⲈ ⲘⲘⲞⲞⲨ´ ⟦ⲚⲦ⟧ⲈⲀⲨⲚⲦⲞⲨ ⲌⲰⲤ ⲘⲀⲢⲦⲨⲢⲞⲤ ⲈⲒ ⲄⲈ
ⲌⲈⲚⲖⲀⲀⲨ ⲀⲚ ⲌⲞⲖⲰⲤ ⲚⲈ

*n-rôme nt-a-u-kanônize mmo-ou e-a-u-nt-ou hôs marturos **ei ge** hen-laau an holôs
ne*

'… the people summoned and presented (*by the deacon John*) to you (*arbiters*)
as if they were witnesses, **if in truth** they aren't'.

The particle ⲄⲈ is not attested in Förster's dictionary of Greek words in Coptic
documentary texts since it is misread in the printed editions of P.Budge, and it is not
attested in *Gertrud Bauer's Zettelkasten Online*, a comprehensive lexicographical
source of non-inflecting Greek words in Coptic through all Coptic dialects and
genres of text.

The phrase **διὰ τούτου** 'therefore' is twice attested in P.Budge, in both cases
used by John:

(32) Line 20 (John):

ⲆⲒⲀ ⲦⲞⲨⲦⲞˋ ⲈⲒ<Ⲉ>ⲚⲀⲄⲈˋ ⲚⲀ⟦ϥ⟧Ⲩ ⲌⲀⲠⲎⲒ
dia touto e-i-anage na-u ha-p-êi
'**Therefore** I (*John*) sue them (*Philemon and wife*) because of this house …'

(33) Line 118 (John):

ⲆⲒⲀ ⲦⲞⲨⲦⲞˋ ⲀⲒϢⲀⳓⲈ ⲀⲨⲰ ⲌⲈⲚⲖⲀⲀⲨ ⲀⲚ ⲚⲈ
dia touto a-i-šače auô henlaau an ne
'**Therefore** I have spoken, and they (*the letters*) are not 'nothing''

It is not in Förster's dictionary; *Gertrud Bauer's Zettelkasten* gives fourteen
instances mostly from Gnostic text of the Nag Hammadi corpus.

The compound conjunction ⲈⲚ ⲞⲒⲤ - *en (h)ois* consisting of the Greek
preposition ἐν and an inflected form of the relative pronoun ὅς — both not
normally borrowed into Coptic — is used once in P.Budge in a speech of
Philemon:

(34) Line 192 (Philemon):

ⲈϢⳢⲈ ⲀⲨⲤⲌⲞⲘⲞⲖⲞⲄⲈⲒˋ ⳢⲈ ⲀⲤⲔⲰ ⲔⲖⲎⲢⲞⲚⲞⲘⲞⲤ ⲀⲤⲘⲞⲨ ¹⁹²| ⲈⲚ ⲞⲒⲢ̣ Ⲁ⟦Ⲥ⟧
ϥⲦⲀⲨⲞˋ ⲠⲢⲀⲚ ⲚⲌⲞⲒⲚⲈˋ ⲚⲎⲦⲚ̄ Ⲛ̄ⲌⲎⲦⲞⲨ

*ešče a-u-homologei če a-s-kô klêronomos a-s-mou ¹⁹²| **en ois** a-f-tauo p-ran n-hoine
nê-tn nhêt-ou*

'If they (*John's party*) admitted that she (*Thekla*) left some heirs and died – **in so far as** he mentioned names of some of them to you – …'

Since this instance of ⲉⲛ ⲟⲓⲥ in P.Budge was not recognised by the first editor, it is not in Förster's dictionary, nor is it attested in *Gertrud Bauer's Zettelkasten*.

The Greek conjunction ἐπάν is used four times by Philemon in semantically and syntactically very similar contexts:

(35) Lines 67–8 (Philemon):

ⲡ̅ϩⲱⲃ ⲟⲩⲟⲛϩ̅ ⲉⲃⲟⲗ **⁶⁸**| ϫⲉ ⲉⲡⲁⲛ ⲁⲥⲧⲁⲁⲩ ⲛⲁⲛ `⟦. . .⟧´ ⲧⲁⲣⲉⲛⲟⲩⲱϩ` ⲛ̅ϩⲏⲧⲟⲩ
ⲙ̅ⲡⲉⲥⲕⲱ` ⲗⲁⲁⲩ ⲉⲡⲱⲥ ⲛ̅ϩⲏⲧⲟⲩ

p-hôb ouonh ebol **⁶⁸**| *če **epan** a-s-taa-u na-n […] tare-n-ouôh nhêt-ou mpe-s-kô laau e-pô-s nhêt-ou*

'It goes without saying that — **since** she gave them (*sc. rooms of her house*) to us that we would dwell in them — she left back nothing of her belongings.'

(36) Lines 82–4 (Philemon):

ⲁⲗⲗ`ⲟⲩⲛ ϥⲙⲡϣⲁ` ⲉⲓⲙⲉ` ϫⲉ ⲉⲡⲁⲛ ⲁⲡⲛⲟⲩⲧⲉ **⁸³**| ⲛⲟϫⲥ̅ ⲉⲡⲉⲧⲛ̅ϩⲏⲧ`
ⲁⲧⲉⲧⲛ̅ⲥⲱⲧⲙ ⲡⲉⲛϩⲱⲃ ϩⲁ ⲡⲛⲟⲩⲧⲉ ⲙⲁⲛⲉⲓ̈ϣⲃⲱ` ⲛⲁⲓ̈ ⲣ̅ ϣⲁⲩ ⲛ̅ⲛ̅ⲧⲟⲩ ⲛ̅ⲛⲁϩⲣⲛ̅
ⲧⲉⲧⲛ̅ⲡⲉⲣⲓⲃⲗ`ᵉⲗᵉ **⁸⁴**| ⲛ̅ϫⲟⲉⲓⲥ

*all'oun f-mpša e-ime če **epan** a-p-noute* **⁸³**| *nočs e-pe-tn-hêt a-tetn-sôtm pe-n-hôb ha-pnoute ma-nei-šbô nai r-šau n-nt-ou nnahrn tetn-perible(ptos)* **⁸⁴**| *n-čoeis*

'But now he is liable to recognise that — **since** God put it in your heart and you listened to our case for God's sake — these stories are not good to be brought before your admirable lords.'

(37) Lines 163–5 (Philemon):

ⲉⲩⲥⲟⲟⲩⲛ **¹⁶⁴**| ⲙⲁⲗⲓⲥⲧⲁ ϫⲉ ⲉⲡⲁⲛ ⲁⲡⲛⲟⲩⲧⲉ ⟦ⲁⲙ⟧†ⲑⲉ ⲁⲙⲙⲁⲧⲉ
ⲛ̅ⲧⲉⲧⲛ̅ⲙⲛ̅ⲧϫⲟⲉⲓⲥ ⲁⲩⲱ ϫⲉ ⲛ̅ⲧⲱⲧⲛ̅ ⲉⲧⲇⲓⲕⲁⲍⲉ ⲉⲣⲟⲛ ϩⲁ ⲡⲛⲟⲩⲧⲉ ⲙⲁ̣ϣⲁϫⲉ
ⲉⲙⲛ̅ **¹⁶⁵**| ⲥⲩⲥⲧⲁⲥⲓⲥ ϩⲓⲱϥ ⲣ̅ ϣⲁⲩ ⲛ̅ⲥⲉⲉⲓⲛⲉ ⲛ̅ⲛⲁϩⲣⲏⲧⲛ̅

e-u-sooun **¹⁶⁴**| *malista če **epan** a-p-noute tⁱ tʰe a-m-mate n-tetn-mntčoeis auô če ntôtn et-dikaze ero-n ha-pnoute ma-šače e-mn-* **¹⁶⁵**| *sustasis hiô-f r-šau n-se-eine nnahrê-tn*

'… especially as they know that — **since** God granted the opportunity and we gained your lordship and it is you who will judge us for God's sake — a word that lacks conclusive force is not worth to be brought before you.'

ἐπάν is once attested outside P.Budge in Förster' *Wörterbuch*, and it is not instanciated in *Gertrud Bauer's Zettelkasten*.

The Greek preposition ἕως 'until' is twice attested in P.Budge, both times used in Philemon's orations and both times co-occuring with the Greek particle γάρ, as a rhetorical means to lead an argument to its extreme point:

> **(38)** Lines 60–1 (Philemon):
> ϩⲉⲱⲥ ⲅⲁⲣ ⲛ̄ⲕⲉⲕⲟⲟϩ` ⲛ̄ **⁶¹|** ⲥⲭⲓⲧⲁⲣⲓⲛ ⲛ̄ⲧⲁⲩⲉⲙⲫⲁⲛⲓⲍⲉ ⲙ̄ⲙⲟⲟⲩ ⲛⲏⲧⲛ̄ ⲉⲩϫⲱ ⲙ̄ⲙⲟⲥ ϫⲉ ⲛ̄ⲧⲁⲥⲉⲡⲓⲧⲣⲉⲡⲉ` ⲛⲁⲩ ϩⲓϣⲟⲩ ⲧⲁⲣⲟⲩⲣ̄ ⲡⲉⲥⲡⲣⲟⲥⲱⲡⲟⲛ
> **heôs** *gar n-ke-kooh n-* **⁶¹|** *skʰitarin nt-a-u-empʰanize mmo-ou nê-tn e-u-čô mmo-s če n-ta-s-epitrepe na-u hiô-ou taro-u-r pe-s-prosôpon*
> '… **even up until** these fragments of drafts which they produced to you, saying that she (*Thekla*) charged them therein to do her legal representation'

> **(39)** Lines 255–6 (Philemon):
> ϩⲉⲱⲥ ⲅⲁⲣ ⲡⲕⲉⲉⲛⲧⲟⲗⲓⲕⲟⲛ **²⁵⁶|** ⲉⲧⲉϣⲁⲣⲉ ⲛ̄ˢⁱᶜⲣⲱⲙⲉ ⲛⲓⲙ ⲥⲙⲛ̄ⲧϥ ⲛⲁ`ⲛ´ⲣⲱⲙⲉ ⲧⲁⲣⲟⲩⲣ̄ ⲡⲟⲩⲡⲣⲟⲥⲱⲡⲟⲛ ⲛ̄ⲥⲉⲉⲛⲁⲅⲉ` ϩⲁⲣⲟⲟⲩ
> **heôs** *gar p-ke-entolikon* **²⁵⁶|** *ete-šare n-rôme nim smnt-f na-n-rôme tar-ou-r p-ou-prosôpon n-se-enage haro-ou*
> '… **even up until** the proxy document that all people would draw up for somebody to do their legal representation and to litigate for them'

ἕως is not otherwise attested in Förster's *Wörterbuch*. The rare attestations of ἕως collected in *Gertrud Bauer's Zettelkasten* make its usage in P.Budge even more conspicuous. *Bauer* encountered ἕως in one single Sahidic text, the famous gnostic treatise called *Pistis Sophia*. As often as it occurs here, a Coptic semantic equivalent co-occurs. If governing a noun, ἕως is paired with the Coptic prepositions ϣⲁ- *ša-* 'until' or ϣⲁ-ϩⲟⲩⲛ ⲉ- *ša-houn e-* 'up until':

> **(40)** Pistis Sophia 10,15.17:
> ϩⲉⲱⲥ ϣⲁ ⲡⲥⲁ ⲛⲃⲟⲗ…ϩⲉⲱⲥ ϣⲁ ⲡⲥⲁ ⲛϩⲟⲩⲛ
> **heôs ša**-*p-sa n-bol* … **heôs ša**-*p-sa n-houn*
> '**until** the outside…**until** the inside'

> **(41)** Pistis Sophia 4,8:
> ϩⲉⲱⲥ ϣⲁϩⲟⲩⲛ ⲉⲡⲟⲟⲩ ⲛϩⲟⲟⲩ
> **heôs ša**-*houn e-poou n-hoou*
> '**up until** the day today'

If governing a clause, ἕως is paired with the Coptic conjugation marker ϢⲀⲚⲦ- *šant-* 'until':

> **(42)** Pistis Sophia 178,23:
> ϨⲈⲰⲤ ϢⲀⲚ†ⲚⲞⲨϨⲘ ⲘⲘⲞⲤ
> ***heôs** šant-ⁱ-nouhm mmo-s*
> '**until** I rescued her'

> **(43)** Pistis Sophia 232,22:
> ϨⲈⲰⲤ ϢⲀⲚⲦⲈⲦⲚ ⲔⲎⲢⲨⲤⲤⲈ Ⲛ̄ϢⲀϪⲈ ⲚⲒⲘ
> ***heôs** šante-tn kêrusse n-šače nim*
> '**until** you proclaimed every word'

Only in a third, different usage — that of a conjunction encoding simultaneity ('as long as') — ἕως occurs without a Coptic semantic synonym, as in one Sahidic and two Manichaean Lycopolitan instances in *Bauer's Zettelkasten*:

> **(44)** NHC VII 114,2:
> ϢⲰϪⲈ ⲠⲚⲞϬ ⲚⲀⲄⲰⲚ ϨⲈⲰⲤ ⲠⲀⲄⲰⲚ ⲔⲀⲀⲦ ⲈϨⲢⲀⲒ
> *šôče p-noc n-agôn **heôs** p-agôn kaat ehrai*
> 'Fight the great fight **as long as** the fight lasts'

> **(45)** Mani Kephalaia 165,21:
> ϨⲈⲰⲤ ⲈⲢⲈⲔⲀⲒⲢⲞⲤ ϢⲞⲞⲠ Ⲁ̄ⲢⲠⲀⲅⲀⲐⲞⲚ
> ***heôs** ere-kairos šoop a-r-p-agatʰon*
> '**as long as** time still exists to do the good'

> **(46)** Mani Kephalaia 227,20:
> ϨⲈⲰⲤ ⲈϤⲦⲎⲔ ⲀⲢⲈⲦϤ ϨⲚⲠⲤⲰⲘⲀ
> ***heôs** e-f-têk aret-f hn-p-sôma*
> '**as long as** he dwells in the body'

P.Budge thus provides rare Coptic evidence of ἕως used as a preposition in its own right, syntactically governing a Coptic noun phrase without support by a Coptic native preposition: ϨⲈⲰⲤ ⲄⲀⲢ Ⲛ̄ⲔⲈⲔⲞⲞϨ *heôs gar n-ke-kooh* '...up until even these fragments', ϨⲈⲰⲤ ⲄⲀⲢ ⲠⲔⲈ ⲈⲚⲦⲞⲖⲒⲔⲞⲚ *heôs gar p-ke-entolikon* '... up until even the proxy document'. The parallelism between the two instances of this usage includes the particle γάρ following ἕως as well as the Coptic quantifier ⲔⲈ

with definite article, thus the same rhetorical twist of 'for even up until'.

The Greek relative ὅπερ is attested twice in P.Budge, both times in Philemon's speech, again in strikingly similar syntactic and semantic environments:

> **(47)** Line 235 (Philemon):
>
> ϣⲁⲥⲁⲡⲁⲛⲧⲁ ⲛⲁⲩ ⲉⲣ̅ ⲡⲉⲧⲉⲡⲱⲛ ⲙ̅⟦ⲱ⟧ⲡⲱⲟⲩ **ⲟⲡⲉⲣ** ²³⁶| ⲁⲕⲣⲓⲃⲱⲥ
> ⲧⲛ̅ⲡⲗⲏⲣⲟⲫ<ⲟⲣ>ⲉⲓⲥⲑⲁⲓ`  ⲉ⟦ ֹ ⟧ⲡⲁⲓ  `ⲭⲉ´ ⲙⲁⲥⲁⲡⲁⲛⲧⲁ ⲛⲁⲩ
>
> *ša-s-apanta na-u e-r-p-ete-pô-n m-pô-ou **oper** ²³⁶| akribôs tn-plêropʰoreistʰai e-pai čе ma-s-apanta na-u*
>
> '(*John's party hopes that*) they will succeed to make ours into theirs — **which is** exactly what we (*Philemon's party*) are convinced they won't succeed —…'

> **(48)** Line 257 (Philemon):
>
> ⲉⲩⲟⲩⲱϣ ⲥⲁⲣⲡⲁⲍⲉ ⲙ̅ⲡⲉⲛⲏⲓ̈ **ⲟⲡⲉⲣ** ⲧⲛ̅ⲡⲓⲥⲧⲉⲩⲉ`  ⲭⲉ  ⲙⲁⲡⲛⲟⲩⲧⲉ  †  ⲑⲉ`
> ⲛ̅ⲥⲉⲙⲁⲧⲉ ⲙ̅ⲡⲁⲓ̈
>
> *e-u-ouôš harpaze m-pe-n-êi **oper** tn-pisteue čе ma-p-noute tⁱ tʰe n-se-mate m-pai*
>
> 'since they (*John's party*) want to steal our house — **what**, we believe, God will not allow them to achieve —…'

Apart from the two instances in P.Budge, the relative ⲟⲡⲉⲣ *oper* is attested in Förster' *Wörterbuch* only in a Greek text (CPR IV 34) and in Greek multi-word expressions (ⲟⲡⲉⲣ ⲙⲏ ⲉⲓⲏ / ⲅⲉⲛⲟⲓⲧⲟ *oper mê eiê / genoito* 'which shall not happen'). *Bauer's Zettelkasten* offers a single instance of a Coptic usage comparable to the P.Budge examples:

> **(49)** Pères apostoliques ed. Lefort, 47,8:
>
> ϣⲗⲏⲗ ⲥⲁⲣⲟⲟⲩ ⲥⲉⲛⲁϣⲙⲉⲧⲁⲛⲟⲓ ⲥⲟⲡⲉⲣ ⲡⲁⲓ ⲙⲟⲕⲥ ⲛⲁⲩ ⲟⲩⲛⲧⲉϥⲧⲉⲝⲟⲩⲥⲓⲁ
> ⲇⲉ ⲉⲡⲁⲓ ⲛ̄ϭⲓ ⲓⲥ ⲡⲉⲭⲥ
>
> *šlêl haro-ou se-na-š-metanoi **hoper** pai mokh na-u ounte-f-t-ekʰousia de e-pai nci I(êsous) pe-Kʰ(risto)s*
>
> 'Pray for them that they will manage to repent — **which** is difficult for them, which however Jesus Christ has power to grant.'

Table 2 shows the distribution of the particles and conjunctions, the preposition and the relative pronoun discussed above. None of them is used by both opponents, the most of them occur exclusively in speeches delivered by Philemon.

Table 2. Borrowed function words: Distribution of hapax legomena and rare words in P.Budge

Lexeme	Philemon	John	Förster, Wb	GBZO
γε – ⲉⲓ ⲅⲉ *ei ge* 'if indeed'	+	-	-	-
διὰ τούτου – ⲆⲓⲀ ⲧⲟⲩⲧⲟ *dia touto* 'therefore'	-	+	-	Some
ἐν / ὅς, ἥ, ὅν – ⲉⲛ ⲟⲓⲥ *en ois* 'in that'	+	-	-	-
ἐπάν – ⲉⲡⲀⲛ *epan* 'since'	+	-	P.Budge+1	-
ἕως – ϩⲉⲱⲥ *heôs* 'until'	+	-	P.Budge	Some
ὅσπερ, ἥπερ, ὅπερ – ⲟⲡⲉⲣ *oper* 'which'	+	-	P.Budge+Greek	Greek+1

5 Discussion and conclusion

Our examination of *manuscript features* (Section 3.2) of the hearing protocol P.Budge led us to the conclusion that the extant item cannot be the original transcript taken during the hearing but must be a clean copy drawn up by the scribe *post hoc*, although one that underwent further corrections and bears the participants' signatures, thus not a copy of the copy.

Our examination of *morphosyntactic features* of P.Budge (Section 4.2) led to the following observations:

1 – The linguistic profile of P.Budge, a Sahidic matrix embedding a number of nonstandard features, is in overall keeping with linguistic repertoires of other Coptic documentary papyri.

2 – Some features hint to the Upper Egyptian background of the participants of the P.Budge hearing, — the disputing parties, the arbiters, and the scribe (examples 7–11).

3 – Exceptional Middle Egyptian language features surfacing in letters from Thekla (and Mena) point to their dwelling place at Middle Egypt (examples 12–14)

Our examination of *lexical features* of P.Budge (Section 4.3) led to the following observations:

4 – P.Budge contains an amazing number of Greek words which are not or rarely attested elsewhere in Coptic. Many (though not all) of them are juridical

terms related to the case of the P.Budge hearing.

5 – P.Budge contains remarkably many Greek function words, significantly beyond the average ratio of Coptic texts (Table 1), some of them rare or even *hapax* items in Coptic, some of them used in different ways — less restricted, more integrated — than elsewhere in Coptic (examples 31–49).

6 – All of those rare Greek function words are not shared by both parties of the P.Budge hearing, but belong to the language of one of them, mostly to that of Philemon (Table 2).

Observations 1 and 2 connect the language of P.Budge to well-known and well-attested linguistic registers of Coptic. Observations 4 and 5 indicate linguistic innovation in the language of P.Budge. According to our *a priori* assumption (Section 4.1), the scribe of P.Budge, when processing the original transcripts into a coherent document, would have widely eliminated the participants' linguistic 'fingerprints' and reshaped their utterances along the line of his own linguistic habits. This, however, is exactly not what the examination and assessment of its linguistic traits tells us. Observations 4 and 5 make it likely to assume, observations 3 and 6 inevitable to conclude that P.Budge preserves individual idioms, since it is not plausible to imagine that its scribe assigned characteristic words and expressions up to function words in the way a novel writer does to the 'main characters' of his text.

Given such evidence for *extensive verbatim quotation*, the question must eventually be raised whether the text of the speeches as conveyed in P.Budge was really (and exclusively) established by oral performance, or whether it was at least partly accessible to the scribe in written records in the same way as Theclas and Menas letters? Our general idea of an arbitration hearing evokes the setting of a physical meeting of the two parties in front of the arbiters with speeches delivered orally. Indeed the meta-language of P.Budge referring to the P.Budge hearing supports this idea for the most part. E.g., the final stipulation of the parties is phrased like this:

ⲀⲚⲞⲔ … ⲚⲀⲒ̈ ⲚⲈ ⲚϢⲀⲬⲈ Ⲛ̄ⲦⲀⲒⲦⲀⲨⲞⲞⲨ` ⲀⲨⲰ ⲘⲚ̄ⲦⲀⲒ ⲔⲈϢⲀⲬⲈ` Ⲙ̄ⲘⲀⲨ ⲈⲬⲰ`
ⲌⲀ ⲦⲈⲒⲨⲠⲞⲐⲈⲤⲒⲤ ⲦⲀⲒ̈

anok … nai ne n-šače nta-i-taouo-ou auô mnta=i ke-šače mmau e-čô ha tei-hypothesis tai

'I …, these are the words I **uttered**, and I have no other word to **say** about this legal cause'

Trying to find arguments *against orality* and to trace instead *signs of written record*, we find indeed two expressions which support a scenario of written submission. First, Philemon says (or rather, his scribe wrote) at the end of his first speech:

(50) P.Budge 95-96

ⲧⲉⲧⲛ̄ⲥⲟⲟⲩⲛ ⲛ̄ⲧⲗⲟⲓⲡⲁⲥ [ⲙⲡⲉⲓ̈]ϩⲱⲃ · ϫⲉ ⲉⲛⲉⲓ̈ⲥϩⲁⲓ̈ ϩⲁϩ ⲛ̄´ϣⲁϫⲉ` [[ⲛ]]ⲧ[[ⲁ]] ⲛ̄ϭⲱ` ⲉ[[ⲓ̈]]ⲛ̄ϯ ϩⲓⲥⲉ` ⲛ̄ⲧⲉⲧⲛ̄ⲙⲛ̄ⲧϫⲟⲉⲓⲥ ·

tetn-sooun n-t-loipas [m-pei-h]ôb če e-n-shai hah n-šače nta-n-cô e-n-ti hise n-tetn-mntčoeis

'you know the rest [of this] case, so that I will not **write** many words and we must not further trouble your lordships.'

Second, at the end of the same speech Philemon says (or rather, his scribe wrote):

(51) P.Budge 109-112

ⲧ`ⲛ̄´ⲡⲁⲣⲁⲕⲁⲗ[ⲉⲓ]  |  ⲟⲩⲛ ⲛ̄ⲧⲉⲧⲛ̄ⲡⲉⲣⲓⲃⲗᵉⲗᵉ ⲛ̄ϫⲟⲉⲓⲥ ϩⲱⲥ ⲉ[[ⲓ̈]]ⲛⲥⲟⲟⲩⲛ ϫⲉ ⲑⲟⲧⲉ ⲙ̄ⲡⲛⲟⲩⲧⲉ ⲟⲩⲏϩ` ⲛ̄ϩⲏⲧ ⲧⲏⲩⲧⲛ̄ ⲁⲩⲱ ⲙⲁⲧⲉⲧⲛ̄ϫⲓ ϩⲟ` ⲛ̄ⲣⲱⲙⲉ ⲛ̄ⲧⲉⲧⲛ̄|ⲣⲟⲉⲓⲥ ˢⁱᶜ ⲡⲇⲓⲕⲁⲓⲟⲛ ⲛⲁ[[ⲓ̈]]ⲛ ϫⲉⲕⲁⲥ ⲉⲣⲉⲡϫⲟⲉⲓⲥ ⲓ̅ⲥ̅ ⲡⲉⲭ̅ⲥ̅ ⲛⲁϩⲁⲣⲉϩ` ⲉⲣⲱⲧⲛ̄ ⲙⲛ̄ ⲛⲉⲧⲛ̄ϣⲏⲣⲉ ⲛ̄ⲟⲩⲛⲟϭ ⲛ̄ⲟⲩⲟⲉⲓϣ ⲛ̄ⲉⲓⲣⲏⲛⲓⲕⲟⲛ | ⲉⲧⲉⲧⲛ̄ϫⲏⲕ ⲉⲃⲟⲗ ⲥⲱⲙⲁ ⲯⲩⲭⲏ ⲡ̅ⲛ̅ⲁ̅ · ✝ ⲇⲉⲥⲡᵒ ✝

tn-parakalei oun n-te-tn-periblᵏʰlᵏ n-čoeis hôs e-n-sooun če t-hote m-p-noute ouêh nhêt-têutn auô ma-tetn-či-ho n-rôme n-tetn-roeis pdikaion na-n čekas ere-p-čoeis I(êsou)s pe-Kʰ(risto)s na-hareh erô-tn mn ne-tn-šêre n-ou-noc n-ouoeiš n-eirênikon e-tetn-čêk ebol sôma pˢukʰê pn(eum)a + despo(tai) +

'We now beseech you most illustrious lordships, knowing that the fear of God dwells within you, and that you are not corruptible by anyone, to observe the right for us, so that (in turn) the Lord Jesus Christ may preserve you and your children for a long, peaceful time and let you be unharmed (in) body, soul and spirit. + Lords! +'

The last word, framed by two crosses, is the abbreviated term + ⲇⲉⲥⲡᵒ + (δέσποτα(ι)), i.e., a Greek vocative form 'Lord(s)'. This very form is found as final acclamation in Greek (and rarely also in Coptic) petitionary letters to high-ranking addressees. It is not likely to be part of an oral performance, as also the two framing crosses, having a graphemic but not a phonetic value, are likely to have been copied from a written *Vorlage*. Also, the preceding wishes for health and well-being have an (epistolo)graphic rather than (rhet)or(ic)al flavor.

Therefore we may conclude that at least the first speech of Philemon was not (only) delivered on the spot, but rather recorded on papyrus and submitted to the arbiters, and eventually copied into the protocol by the P.Budge scribe as he copied also the letters of Thecla and Mena.

Eventually, it is striking to find the speech of the peasant Philemon so much loaded with *innovative* language (observations 4, 5, 6). Wondering how an illiterate peasant was capable to perform so well and inventively as an orator, one may remember Philemon's own reference to people 'who do have knowledge' (Section 2, example 5) and have given him advise. His access to such people, though, raises another, no less vexing question — that of his social status, as does his economic power that allowed him first to grant a loan of money to a destitute widow, and then to pay off the *iustum pretium* with the multiple amount of money. Even our initial question (Section 1) why and how the hearing protocol of P.Budge could arrive at Philemon's private archive may be somehow related to his personal status and connections. But this is another story which lies beyond the scope of this study.

References

Allam, Shafik 1991. 'Glossen zu einem schiedsrichterlichen Verfahren (kopt. p.Budge + griech. pBM 2017)', *Mitteilungen des Deutschen Archäologischen Instituts Kairo* 47: 1–9.

Allam, Shafik 1992. 'Observations on Civil Jurisdiction in Late Byzantine and Early Arabic Egypt', in J. Johnson (ed.), *Life in a Multicultural Society: Egypt from Cambyses to Constantine and Beyond*, Studies in Ancient Oriental Civilization 51. Chigaco, 1–8.

Bacot, Seÿna 2009. *Ostraca grecs et coptes des fouilles franco-polonaises sur le site de Tell Edfou (O.EdfouCopte 1–145)*, Bibliothèque d'études coptes 19, Le Caire: IFAO.

Bagnall, Roger S. 2011. 'Greek and Coptic in Late Antique Egypt', in Roger S. Bagnall, *Everyday Writing in the Graeco-Roman East.* Sather Classical Lectures 69. Berkeley – Los Angeles – London: UCP, pp. 75–94.

Bagnall, Roger S. and Worp, Klaas A. 2004. 'Dating the Coptic legal documents from Aphrodite'. *ZPE* 148: 247–52.

Berkes, Lajos and Vanthiegem, Naijm 2019. 'A late Coptic Tax Receipt From the Egyptian National Library (P.Cair. Nat. Library inv. 3890)', *Journal of Coptic Studies* 21: 13–17.

Bosson, Nathalie 2006. 'Remarques sur la « structure (ϩ)ⲁ- … (ϩ)ⲁϥ- »', *Lingua Aegyptia* 14: 281–300.

Boud'hors, Anne and Shisha-Halevy, Ariel 2012. 'Two remarkable features of Coptic syntax', *Zeitschrift für Ägyptische Sprache und Altertumskunde* 139: 105–12.

Boud'hors, Anne, Delattre, Alain et al. 2018. 'Un nouveau départ pour les archives de Papas. Papyrus coptes et grecs de la jarre d'Edfou', *BIFAO* 117: 87–124.

Boulard, Louis 1913. 'La vente dans les actes coptes', *Études d'histoire juridique offertes à Paul Frédéric Girard*, tome II. Paris, pp. 1–94.

Crum, Walter E. 1925. 'Koptische Zünfte und das Pfeffermonopol', *Zeitschrift für Ägyptische Sprache und Altertumskunde* 60: 103–11.

Crum, Walter E. 1926. 'The language of the texts', in Walter E. Crum & Herbert E. Winlock, *The Monastery of Epiphanius at Thebes.* The Metropolitan Museum of Art Egyptian Expedition, part 1, New York 1926, pp. 232–56.

Delattre, Alain and Fournet, Jean-Luc 2011. 'Les ostraca grecs et coptes d'Edfou. À propos d'une publication récente', *Archiv für Papyrusforschung und verwandte Gebiete* 57: 79–98.

Depauw, Mark 2003. 'Autograph confirmation in Demotic private contracts', *Chronique d'Egypte* 78: 66–111.

Foss, Clive 2009. 'Egypt under Muʿāwiya. Part I: Flavius Papas and Upper Egypt', *Bulletin of the School of Oriental and African Studies* 72: 1–24.

Fournet, Jean-Luc 2010. 'Sur les premiers documents juridiques coptes', *Études coptes XI. Treizième journée d'études (Marseillle, 7–9 juin 2007)*. Paris, pp. 125–37.

Fournet, Jean-Luc 2019. *The Rise of Coptic. Egyptian versus Greek in Late Antiquity*. Princeton & Oxford: Princeton UP.

Förster, Hans, Fournet, Jean-Luc and Richter, Tonio S. 2012. 'Une *misthôsis* copte d'Aphrodité (P.Lond. inv. 2849): le plus ancien acte notarié en copte?', *Archiv für Papyrusforschung* 58: 344–59.

Frantz-Murphy, G. 1993. 'Settlement of property disputes in provincial Egypt: The reinstitution of courts in the Early Islamic Period', *Al-Masâq. Studia Arabo-Islamica Mediterranea* 6: 95–105.

Frantz-Murphy, G. 2003. 'The reinstitution of courts in Early Islamic Egypt', *Bulletin de la Société Archéologique d'Alexandrie* 47: 71–85.

Gascou, Jean and Worp, K. A. 1982. 'Problèmes de documentation apollinopolite' *ZPE* 49: 83–95.

Gertrud Bauer Zettelkasten Online <http://research.uni-leipzig.de/ddglc/bauer/>

Grossman, E. 2009. 'Protatic ⲉϥⲥⲱⲧⲙ revisited', in A. Giewekemeyer, G.Moers & K.Widmaier (eds.), *Liber amicorum - Jürgen Horn zum Dank*. Göttinger Miszellen Beihefte 5, Göttingen: pp. 47–56.

Grossman, Eltan (unpubl.). 'Breaking into the noun phrase: The grammaticalization of a focalized possessor construction in Coptic', unpublished working paper.

Grossman, Eltan and Haspelmath, Martin 2015. 'The Leipzig-Jerusalem Transliteration of Coptic', in Eltan Grossman, Martin Haspelmath and Tonio S. Richter (eds.), *Egyptian-Coptic Linguistics in Typological Perspective*, Empirical Approaches to Language Typology 55, Berlin – Munich – Boston: De Gruyter Mouton: pp. 145–53.

Hasitzka, Monica R.M. 1993. *Koptisches Sammelbuch*, Bd. I. *MPER* n.s. XXIII/1, Wien: Brüder Hollinek.

Kahle, Paul E. 1954. *Bala'izah. Coptic Texts from Deir el-Bala'izah in Upper Egypt*. Oxford – London: Oxford University Press.

Kreuzsaler, Claudia 2010. 'Die Beurkundung außergerichtlicher Streitbeilegung in den ägyptischen Papyri', in C. Gastgeber (Hsg.), *Quellen zur byzantinischen Rechtspraxis. Aspekte der Textüberlieferung, Paläographie und Diplomatik. Akten des internationalen Symposiums Wien, 5. –7. 11. 2007*. ÖAW, phil.-hist. Kl., Denkschriften 413. Wien 2010, 17–26.

Layton, Bentley 2011. *A Coptic Grammar: With Chrestomathy and Glossary: Sahidic Dialect*. Porta linguarum orientalium n.s. 20. Wiesbaden 2011³.

Lewis, Naphtali 1993. 'The demise of the Demotic document: when and why?', *Journal of Egyptian Archaeology* 79: 276–81.

Ludwig, M. 2009. Übersetzung und Studien zum Anhörungsprotokoll des *Pap. Budge (P.Col. Inv. 600 [SBKopt I 036])*. Unpublished Master's ThesisUnpublizierte Magisterarbeit, Universität Leipzig.

MacCoull, Leslie S. B. 2007. 'Why do we have Coptic documentary papyri before A.D. 641?', in N. Bosson and A. Boud'hors (eds.), *Actes du huitième congrès international d'études coptes,* OLA 163. Leuven, vol. 2: 751–8.

MacCoull, Leslie S. B 2013. 'Niches in an Ecosystem: The Choice of Coptic for Legal Instruments in Late Antique Egypt', *Analecta Papyrologica* 25: 257–76.

Muhs, Brian 2005. 'The *grapheion* and the disappearance of Demotic contracts in early Roman Tebtynis and Soknopaiou Nesos', in S. Lippert and M. Schentuleit (eds.), *Tebtynis und Soknopaiou Nesos – Leben im römischen Fayum*, Wiesbaden: pp. 93–104.

Nagel, Peter 1991. Entry 'Akhmimic', in *Coptic Encyclopedia*, vol. 8. pp. 19–27

Rémondon, R. 1953. *Papyrus grecs d'Apollônos Anô.* DFIFAO 19. Le Caire.

Richter, Tonio S. 2008. *Rechtssemantik und forensische Rhetorik.* 2nd ed.*,* Philippika. Marburger altertumskundliche Abhandlungen 20. Wiesbaden 2008.

Richter, Tonio S. 2010. 'Koptische Rechtsurkunden als Quellen der Rechtspraxis im byzantinischen und frühislamischen Ägypten', in C. Gastgeber, F. Mitthof and B. Palme. (eds.), *Quellen zur byzantinischen Rechtspraxis Aspekte der Textüberlieferung, Paläographie und Diplomatik. Akten des Symposiums Wien, 5–.7. November 2007*, Wien 2010: 39–59.

Richter, Tonio S. 2013. 'O.BM EA 20300. In search for the latest dated Demotic ostracon', in Rodney Ast, Hélène Cuvigny, Todd M. Hickey, and Julia Lougovaya (eds..), *Papyrological Texts in Honor of Roger S. Bagnall.* American Studies in Papyrology 53: pp. 285–90.

Richter, Tonio S. 2019. 'Die koptischen Papyri aus dem „Ankauf aus Edfû (Apollinopolis magna) vom Jahre 1911" der Jenaer Papyrussammlung (P.Ien. Inv. 446–53)', in *Archiv für Papyrusforschung* 65: 156–75.

Richter, Tonio S. 2020. 'P.Berlin P.8316 (= BKU I, 21), ein koptisches Rezept zur Purpur-Imitation durch Krapp-Färbung auf gebeizter Wolle', *Journal of Coptic Studies* 22: 151–86.

Schiller, A. A. 1964 'The Interrelation of Coptic and Greek Papyri: P.Bu. and P.BM inv. nos. 2017 and 2018', in *Studien zur Papyrologie und antiken Wirtschaftsgeschichte. Festschrift Friedrich Oertel*. Bonn. (Zusammenführung

der Londoner Teile des Archivs des Philemon mit P.Col. inv. 600), pp. 107–19

Schiller, A. Arthur 1968. 'The Budge Papyrus of Columbia University', *Journal of the American Research Centre in Egypt* 7: 79–118 (*ed.princ.* of P.Col. inv. 600, reedited in *SBKopt.* I 036).

Schiller, A. Arthur 1971. 'The Courts Are No More', in *Studi in onore di Edoardo Volterra,* vol. I, Milano, pp. 469–502.

Shisha-Halevy, Ariel 1973. 'Apodotic ечсштм: A hitherto Unnoticed Late Coptic, Tripartite Pattern Conjugation-Form and its Diachronic Perspective', *Le Muséon* 86: 455–66.

Shisha-Halevy, Ariel 1974. 'Protatic ечсштм: a hitherto unnoticed Coptic tripartite conjufation form and its diachronic connections'. *Orientalia* 43: 369–81.

Shisha-Halevy, Ariel 1976. 'Akhmîmoid Features in Shenute's Idiom', *Le Museon* 89: 353–66.

Shisha-Halevy, Ariel 1988. *Coptic Grammatical Chrestomathy*. Orientalia Lovaniensia Analecta 30. Leuven: Peeters.

Shisha-Halevy, Ariel 1991. 'Entry Sahidic', in *Coptic Encyclopedia* vol. 8, pp. 195–202.

Stadler, Martin A. 2008. 'On the Demise of Egyptian Writing: Working with a Problematic Source Basis', in J. Baines, J. Bennet, and S. Houston (eds.), *The Disappearance of Writing Systems. Perspectives on Literacy and Communication*. London – Oakville: Equinox, pp. 157–8.

Steinwenter, Arthur 1920. *Studien zu den koptischen Rechtsurkunden aus Oberägypten*. Studien zur Palaeographie und Papyrusforschung XIX, Leipzig.

Till, Walter 1928. *Achmîmisch-koptische Grammatik mit Chrestomathie und Wörterbuch*, Leipzig: Hinrichs.

Zauzich, Karl-Theodor 1983. ‚Demotische Texte römischer Zeit‘, in G. Grimm, H.Heinen, E. Winter (Hsgg.), *Das römisch-byzantinische* Äypten, Ægyptiaca Treverensia II, Mainz: pp. 77–80.

Zilliacus, Henrik 1940 'Griechische Papyrusurkunden des VII. Jahrhunderts n.Chr.‘, *Eranos* 38: 79–107.

10. Early Medieval Scribes' Command of Latin Spelling and Grammar: A Quantitative Approach

Timo Korkiakangas

1 Background, objectives and hypotheses

This Chapter investigates the association between spelling variation and certain classical Latin grammatical categories in early medieval documents of Tuscany. The investigation will include linguistic, sociolinguistic, cognitive and educational aspects. The Latin of the Tuscan documents of the eighth and ninth centuries is often highly nonstandard in terms of both grammar and spelling. Although documentary Latin can be described as a language variety of its own with several special terms and fixed expressions (Bartoli Langeli 2006: 25ff.; Larson 2012: 66; cf. Herman 1992: 185), it can also be justly compared to the classical Latin spelling and morphological norm, which the scribes still seemed to pursue to varying extents and which some scribes managed quite well (Korkiakangas 2016: 36; Amsler 1993). The actual models were the Bible and the Christian authors of late antiquity whose Latin can be described post-classical in stylistic terms but is essentially classical in regard to grammar. Classical and standard Latin will be used as synonyms in the following Sections.

Leaving aside the few tentative attempts at recording the vernacular in writing that have come down to us (Castellani 1976), classical Latin was the only language variety with an established writing system that was available. Written Italian vernaculars only came into broad use later in the Middle Ages. Thus, the considerable linguistic variation found in the documents results from the fact that the spoken language (the scribes' L1) and written Latin (L2 to be learnt) had been deviating for a significant period, and the documentary scribes allowed their L1 to interfere with the L2 to varying degrees. In general, it can be assumed that the degree to which scribes produced classical Latin forms and constructions reflected their learned language skills, i.e. the mastery that scribes had acquired in the L2 under their education.

The objective of this Chapter is to examine quantitatively whether the variation in spelling and the variation in employing grammatical categories are associated with each other. In other words, do skilled spellers utilise classical grammar better than do less skilled ones? My working hypothesis is that the more

nonstandardly a scribe spelled, the more frequently he produced novel Romance-type forms and constructions (innovative L1 features) and the poorer was his command of those classical Latin forms and constructions that had vanished or were in decline in late Latin (conservative L2 features). This hypothesis will be statistically tested in Section 5. As mentioned previously, the innovative features had crept into the written code from the spoken idiom of the time, while the conservative features were derived from centuries-old legal Latin. The situation was likely to differ fundamentally from modern-language acquisition situations in which oral usage is also learnt, unlike with documentary Latin.

The conservative grammatical features were memorised during L2 acquisition in school and/or by immersion in authoritative model texts written in standard Latin. The same applies to spelling, which is likely to have occupied an important position in instruction (de Paolis 2010: 229–91; Law 1982: 40–41; cf. Cribiore 2001: 186). My working hypothesis assumes that the scribes mastered the conservative school-learned, memorisable features more or less as a bundle and, consequently, these features pattern with each other and differ from the innovative spoken-language features. Here, the particular linguistic situation in early medieval Italy has to be taken into due consideration: standard Latin was exclusively a medium of written communication and was not meant to be spoken, while all oral communication took place in the vernacular variety, which must have resembled Old Italian to a significant degree (Herman 1991: 39–41; Walsh 1991: 205–7; Zamboni 2000: 182; Lüdtke 1964). Only by the Carolingian ecclesiastical reforms at the end of the eighth century did the emperor require that liturgy and prayers were recited in a restored Latin. However, we do not know what this meant in practice, or what consequences the imperial order had in the kingdom of Italy (Bartoli Langeli 2006: 26–8). It is also often stated that the conceptual differentiation between spoken vernacular and written Latin took place in Gaul in the early ninth century as a result of the Carolingian reforms (e.g. Wright 1991: 109). In Italy, this conceptual rupture is likely to have happened later, given that the first substantial vernacular texts only appear from the tenth or eleventh century onwards (Castellani 1976).

Since the scribes were not and did not seek to become speakers of the L2 they wrote, the mastery of both L2 spelling and grammar was the result of learning; thus, it is justifiable to assume that the spelling- and grammar-related features pattern in a similar way. By contrast, the vast majority of modern-language learners also learn to speak the language, and oral communication is often the most important motivation to learn a language. Consequently, some fluent L2 speakers of modern English, for example, may have invested much less effort in learning

its correct spelling. Thus, for spoken modern languages in which the grapheme/ phoneme correspondence is irregular, the connection between grammatical and spelling competence is not likely to be as close as it was for written early medieval Tuscan Latin, in which both types of competences served written communication exclusively while vernacular was used for oral communication.

It is true that the grammatical systems of Latin and Old Italian were not very different from each other; thus, L2 acquisition was not necessarily a particularly difficult task. On the other hand, the distance between the two languages may have appeared smaller than it actually was, which may have led to indifference to and a failed recognition of the differences that existed in reality. This non-recognition is also likely to have been the reason for the long-standing conceptual non-differentiation between Latin and the vernacular in Italy.

2 Data

Early medieval documents can be considered a more useful resource for variationist linguistic analysis than can most literary texts. As most are originals, the documents have not been subjected to emendations and modifications during textual transmission, unlike texts that have survived in copied manuscripts. The documents also allow addressing sociolinguistically inspired research questions because the writers' names, the places of writing and the dates of almost all the documents are known. Early medieval Italian documents show notable linguistic and formula variation. Along with the fact that no contemporary formulary book is known from Italy, this variation suggests that early medieval Italian documents were not copied slavishly from model document collections, but instead the scribes memorised the legally required formulae, which they then applied and modified as necessary (Schiaparelli 1933: 3). It is easy to see how such a practice resulted in variation at all levels of language, notwithstanding the underlying formulaicity. Besides, standard formulae could not be used to express those details that varied from one legal transaction to another. The scribes had to improvise the case-specific contents on their own, often drawing on the spoken language of the time (Sabatini 1965).

The data of this study is the *Late Latin Charter Treebank* (LLCT). LLCT is a lemmatised and morphologically and syntactically annotated corpus which consists of 519 hand-written private documents (225,834 tokens). They record the selling and buying of landed property and were written by 176 scribes in Lombard and Carolingian Tuscany (mainly Lucca, Siena and Pisa) between 714

and 869 CE. The documents survive as originals, with the exception of a few contemporary copies. The annotation follows the *Guidelines for the Syntactic Annotation of Latin Treebanks* (Bamman *et al.* 2007), with the additions and modifications described in Korkiakangas and Passarotti (2011). The LLCT is based on three copyright-free diplomatic editions and is available for download in Prague Markup Language format (PML) (see References).

3 Spelling variation

Spelling correctness is utilised widely in modern-language studies (e.g. Burt 2006). As for Latin, the rarity of coherent text corpora with original spelling has constrained this kind of research until recently. This Chapter exploits the method of quantifying spelling variation developed in Korkiakangas (2017). I quantify spelling variation as a continuous variable that indicates the percentage of all nonstandardly spelled characters in all characters written by a scribe (e.g. 187 nonstandard spellings in 1,306 characters of the document *ChLA* 732 corresponds to a spelling correctness of 85.7%).

First, the nonstandard word forms of LLCT are normalised into respective standard-Latin forms. An edit distance value is then calculated between all the attested word forms and their standard-Latin counterparts, regardless of whether the attested ones were originally spelled standardly or nonstandardly. Edit distance is the minimum number of single-character changes required to transform one text string into another. I utilise the Levenshtein edit distance where these changes can be insertions, deletions or substitutions. For example, the Levenshtein distance for the attested form *abi* and the respective standard Latin form *habes* 'you have' is three because the two forms differ from each other by three single-character changes (*h* > Ø, *e* > *i*, *s* > Ø). Since the standard form has five characters, the spelling correctness percentage of the attested form *abi* is 40 (two in five). This percentage of correct spelling is then averaged for all the characters found in the documents written by each of the 176 LLCT scribes.

The normalised word forms were acquired by the lemmatizing and morphological tagging of the Open Office Latin lexicon using Whitaker's WORDS tagger (Words 1.97FC). This procedure yielded a two-million-item list of classical Latin word form/lemma/morphological tag entries. I then matched the lemma/morphological tag pair of each LLCT word form with this entry list and obtained the respective classical word forms as output. When calculating the edit distance, the letters *u* and *v*, *i* and *j*, as well as *k* and *c,* were counted

as equivalents because they are merely alternative graphical representations of the same graphemes. Fragmentary and abbreviated words were excluded from the analysis, provided that their inflectional endings could not be restored and expanded with certainty. I explain the procedure in detail in Korkiakangas (2017).

From a linguistic point of view, the spelling deviations of documentary Latin are caused by various motivations. Apart from purely typographical errors, most nonstandard spellings are related to the substantial phonological change that had taken place. However, this phonological change was often tightly connected to morphological (and even syntactic) evolution. For example, the prepositional phrase *ad uscio,* instead of the standard *ad ostium* 'by the door', involves phonologically motivated changes, such as the closure of the stressed /o/ as /u/, the palatalization of *sti* as *sci* [ʃːi], and the reduction of the original inflectional morpheme *-um* as *-o*. This reduction can be interpreted as resulting from the concomitant phonological merger of the syllables in question and from the levelling in prepositional complementation. The latter means that no distinction was made between the originally accusative- and ablative-form complements of prepositions (*-um* and *-o*, respectively) in the spoken language of the time, although the written language usually still cherished that difference. Thus, even spelling indirectly reflects the scribes' knowledge of standard Latin grammar. Since spelling was bound to reflect the drastic linguistic changes, correct spelling was one of the main concerns of school instruction from late antiquity onwards. Everything suggests that classical spelling was considered an integral part of a good competence in Latin.

4 Examined features

In Section 5, the association between spelling variation and a selection of linguistic features will be studied. Those features were selected so that they are indicative of language change and possibly represent different areas of language. In this Section, I present this selection: nine linguistic features that are either conservative or innovative, plus the date. The features are listed below in a tentative order of increasing grammatical complexity: first, lexicon, then, morphology and, lastly, syntax. The complexity of these grammatical domains will be discussed in Section 6.2. Five of the features that will be examined here were already discussed in my Chapter about formulaicity in documentary Latin (Korkiakangas 2018), albeit with different emphasis and operationalization. In the following, I explain the linguistic motivation for each feature and how the feature is operationalised as a variable.

Date. The date variable is included as a reminder that, although the LLCT is analysed here as an essentially synchronic data set, diachronic variation is also present, as I show in Korkiakangas (2017), in which I discuss the historical reasons for this. The continuous variable indicates the middle of each scribe's period of activity based on the LLCT evidence. If only one document survives from a scribe, the scribe's activity period consists of that year. If two or more documents were written by the same scribe in different years, the scribe's position on the timeline is the mean of the first and last dates. Only a few scribes wrote documents over several decades.

Innovative lexicon. The percentage of non-classical lemmas in the total number of words is expected to measure how readily the scribe utilised innovative lexicon. Seventy-nine lemmas were found that are indisputably non-classical: *aldia, aldiaricus, aldio, aldionalis, arimannus, arra, banda, barba, barbanus, batto, bluto, bullitanus, calderarius, cambiator, cambium, caminata, canavarius, cavallarius, cavallicatura, cergiolitum, caesa, concambiatio, concambio, concambium, debluto, fiuwadia, focacia, fossata, fossatum, fumarius, gahagium, gasindus, gastaldus, grunda, launechild, mallo, marepas, marscalcus, monto, morgingabum, mustariolum, ornile, patrinius, paupertacula, petia, petiola, petiolum, petium, rasula, scafilus, scherpha, scufia, sculdahis, spanga, sporus, staffilus, stanclio, stantarium, strata, summarra, sundrialis, sundrium, tessero, tia, tingatio, torta, ubiscarius, usitile, wadia, wadio, waldemannus, waldus, vassus, weregeldum, viccia, viganatio, viganio, vitellata, zapa*. The chronology of the lemmas was checked using several dictionaries. Lexicon is, of course, closely related to the propositional contents of text; therefore, the percentage of the Innovative lexicon is likely to vary from document to document depending on the matter discussed. Due to the skewness of its distribution, the variable was not encoded as continuous. Instead, I divided it in three categories that are roughly equal in size (no innovative lemmas, 0.02% to 0.40% in all words and 0.41% to 3.75% in all words). The variable varies from 0% to 3.75%.

Genitive plural form. The genitive plural form (*-orum/arum/um*) is likely to have decreased in use at a relatively early stage, unlike the genitive singular morphology, as is witnessed by its hypercorrect and crystallised attestations in late nonstandard texts. The form is assumed to have been utilised as a prestige form (Sornicola 2012: 59; Politzer and Politzer 1953: 26, 28). A prestige form is a variant to which language users assign social prestige (Sairio and Palander-Collin 2012). Within the LLCT, prestige variants tend to be classical Latin forms and constructions. The continuous variable indicates the proportion of genitive plural forms in the total number of words in the LLCT (range 0% to 2.4%).

Neuter in *-um* instead of *-o/-u/-us*. The neuter gender had mainly merged with the masculine by the late Latin period (Adams 2013, 383ff.). However, some scribes had memorised a greater or lesser set of neuters, apparently due to instruction in school. Therefore, the degree to which the *o*-stem neuters in the LLCT maintained their conservative nominative/accusative singular *-um* instead of *-o, -u* or *-us* can be used as a hint about a scribe's learned language skills (Korkiakangas 2010: 139; Löfstedt 1961: 226). The continuous variable represents the share of the *-um* ending of *o*-stem neuters in all singular *o*-stem neuters in the LLCT and varies from 0% to 100%.

***De* PP instead of the genitive**. By the early Middle Ages, the prepositional phrase with *de* had gradually replaced the genitive case as the main marker of adnominal possession (Valentini 2018; Fiorentino 1994: 45–6; Adams 2013: 267–74). The genitive morphology still survived, at least in the singular, where it probably merged with the dative singular morphology (Sornicola 2012: 73–4; de Dardel and Gaeng 1992: 99, 103–4; Zamboni 2000: 110). The continuous variable measures the share of the *de* PP in all cases of adnominal possession in the LLCT and varies from 0% to 66.7%.

Complementiser clauses instead of accusative and infinitive construction (ACI). Classical Latin encoded complement clauses of semantic classes, such as those reporting indirect speech, with ACI. The competing complementation strategy, i.e. finite clauses headed by complementisers, such as *ut* or *quod* 'that', became increasingly frequent beginning from the later Empire (Greco 2017; Zamboni 2000: 119–20; Ledgeway 2012: 244ff.). ACI seems to have been considered a conservative prestige variant and continued to be far more common (86%) in the LLCT than is the innovative complementiser clause (14%). The dichotomous variable indicates whether a scribe utilised complementiser clauses in which ACI could have been utilised or not.

Subjects/direct objects in *-bus*. In the LLCT, the original dative/ablative plural ending *-bus* appears to have been used in several hypercorrect ways with a restricted set of lemmas. For example, the dative/ablative *heredibus* acts as the subject of a finite verb in the example (1) (Fiorentino 1994: 39; Sornicola 2012: 57–8; Politzer and Politzer 1953: 28–9).

> (1) *ChLA* 74.26 (817 CE)
> nos et nostris heredibus tibi et successoribus tuis exinde iustitia per singulos annos reddere debeamus
> we and our heirs must pay you and your successors the lease thereof every year

The origin of these hypercorrect uses as subject and direct object arguments in the LLCT is likely to be found in formulaic prepositional phrases of the type *cum nostris heredibus* 'with our heirs' (Korkiakangas 2016: 48–52). The dichotomous variable indicates whether a scribe utilised *-bus* ending in the subject or direct object function or not.

Accusative subjects. The morphosyntactic alignment of late Latin underwent a partial reorganization, which finally led to the disappearance of the case system as the marker of subject and object arguments in most Romance vernaculars. The accusative, originally the marker of the direct object, is likely to have extended first to those subjects of intransitive finite verbs that underwent verbal action, i.e. were Patients by their semantic role. These, like all subjects, had been originally marked with the nominative. The accusative was later extended to most other subjects as well, thus abolishing the case opposition. In fact, the vast majority of Romance nouns are derived etymologically from the Latin accusative forms (Rovai 2012: 103ff.; Cennamo 2009: 315ff.; Ledgeway 2012: 328ff.). LLCT is practically the only Latin corpus with copious occurrences of the accusative in the subject function (Korkiakangas 2016). In the early medieval Latin of Italy, several word-final sounds had weakened, which led to ambiguity regarding the nominative and accusative inflections in almost all the declensional classes. Therefore, only those non-pronominal subjects in which the nominative and accusative forms remained clearly distinguishable due to the different number of syllables are included in the analysis (e.g. *vir-tus*.NOM as opposed to *vir-tu-te*(*m*).ACC 'virtue, church'; Korkiakangas 2016: 111). The continuous variable measures the proportion of accusative subjects in all the subjects of finite verbs in the LLCT and varies from 0% to 100%.

OV word order. The transformation of the pragmatically conditioned word order in classical Latin to the syntactically conditioned one of the Romance languages is known to have been slow. The most typical verb/direct object order in classical Latin was OV, while VO dominates in Romance languages, albeit with considerable variation (Spevak 2010; Ledgeway 2012: 225–35). The transformation was likely to have been in progress at the time of the LLCT, with OV still being the most frequent order (64.7%). The OV order is included in the present analysis because it has sometimes been argued that it was considered a conservative prestige feature from a very early stage (see the references in Ledgeway 2012: 229). I decided to examine only the most prototypical and unambiguous verb/object combinations with one non-coordinated finite verb and non-pronominal direct object (Korkiakangas 2016: 196; Korkiakangas 2018: 437). The continuous variable measures the

proportion of OV-ordered clauses in all the LLCT clauses described above and varies from 0% to 100%.

Past participle/auxiliary verb order. The preferred classical order for the auxiliary verb *sum* 'to be' and the past participle in the perfective forms of passive and deponent verbs is past participle/auxiliary. This classical order shows a decline in the history of Latin in favour of the auxiliary/past participle order, which is typical of modern Romance languages (Bauer 2006: 294; Danckaert 2016; Ledgeway 2012: 223–4). The conservative past participle/auxiliary order is still the prevailing order in the LLCT (90.0% of occurrences), but the innovative order is seen at least once for 36% of the scribes. The dichotomous variable indicates whether a scribe utilised the past participle/auxiliary order in all past participle/auxiliary combinations or not.

Five of the features above are encoded as continuous variables (Date, Genitive plural form, Neuter in *-um*, *De* PP, OV word order) and five as categorical variables (Innovative lexicon, Complementiser clauses, Subjects/objects in *-bus*, Accusative subjects, Past participle/auxiliary verb order), the latter four of which are dichotomous. The features that are here encoded as categorical variables with two or three categories would have particularly skewed distributions (over 30% of occurrences with the same value, which is either 0% or 100%) if they were treated as continuous variables.

5 Statistical analysis

The statistical association of correct spelling and the features presented in the previous Section is examined below using the general linear model procedure (Carey 2013: 129ff.). The general linear model was chosen because it can measure the effects of both continuous and categorical independent (explaining) variables on the means of various groupings of a continuous dependent variable (the variable to be explained); here, the dependent variable is the correctness of the spelling. It is customary to first present the bivariate analysis of the associations between the dependent variable and each independent variable individually. The F-statistic indicates the probability that the group means are equal (Carey 2013: 115–16). The higher the value of the F-statistic, the more likely the association in question is statistically significant. The core of the statistical analysis is the full model, which analyses the effect of all the independent variables on the dependent variable. The coefficient β indicates how the value of the dependent variable changes when an independent variable varies, while the other independent variables remain

fixed. A positive β predicts an increase in the dependent variable, and a negative β predicts a decrease. The greater the absolute value of the β, the stronger the prediction. N is the number of occurrences (here scribes) and s.e. is the standard error. Note that the β coefficients of categorical variables are compared to a conventional baseline category marked 'a'. With continuous variables, the change in β is compared to a one-unit change on the variable's scale (percentages).

Table 1. The effect of ten variables (see Section 4) on spelling correctness.

Independent variables		Bivariate analysis		Full model	
	N	β	s.e.	β	s.e.
Date (F)		59.38***			
year, scribe's average	176	0.0003***	$3.8*10^{-5}$	-0.0003***	$3.6*10^{-5}$
Innovative lexicon (F)		3.97*			
0%	57	-0.02 (n.s.)	0.004	0.001 (n.s.)	0.003
0.02% to 0.40% in all words	60	0.009*	0.004	0.007*	0.003
0.41% to 3.75% in all words	59	a		a	
Genitive plural form (F)		7.72**			
% proportion of gen. pl. in all words	176	0.873**	0.314	0.944***	0.266
Neuter in -*um* instead of -*o*/-*u*/-*us* (F)		6.62*			
% share of -*um*	159	0.013*	0.005	0.012**	0.004
***De* PPs instead of genitive (F)**		11.38***			
% share of *de* PP	176	-0.042	0.012	-0.051***	0.01
Compl. clause instead of ACI (F)		16.44***			
no	88	-0.013***	0.003	-0.007**	0.003
yes	88	a		a	

Subjects/objects in *-bus* (F)		9.31**			
no	106	-0.010**	0.003	-0.001 (n.s.)	0.003
yes	70	a		a	
Accusative subjects (F)		0.344 (n.s.)			
no	129	0.002 (n.s.)	0.004	0.002 (n.s.)	0.003
yes	47	a		a	
OV word order (F)		2.26 (n.s.)			
% share of OV order	173	0.009 (n.s.)	0.006	0.007 (n.s.)	0.005
Past participle/auxiliary verb order (F)		0.432 (n.s.)			
not all PPart/auxiliary	64	0.002 (n.s.)	0.003	-0.001 (n.s.)	0.003
all PPart/auxiliary	112	a		a	
Adjusted R^2 (full model)				0.458	
*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, n.s. = not significant, s.e. = standard error, a = baseline category					

All associations that are statistically significant (see asterisks) in the bivariate analysis, except for the Subjects/objects in *-bus*, are also statistically significant in the full model. Accusative subjects, OV word order and Past participle/auxiliary verb order did not show statistically significant associations in either analysis. The following list summarises the interpretations of the statistically significant associations in the full model:

1) The more advanced the **Date** within the time frame under examination, the smaller the degree of correct spelling. The decreasing effect is, however, slight, and the bivariate analysis suggests an opposite trend: spelling correctness increases slightly by time if the effect of the other independent variables is not taken into account. The fact that this kind of diachronic change is involved is a reminder

that all the variations present in the LLCT data are by no means explained by synchronic linguistic features alone.

2) The category of 0.02% to 0.40% of **Innovative lexicon** is associated with greater spelling correctness than is the category of 0.41% to 3.75%. Even though the small coefficient β of the category 'no Innovative lexicon' is positive, it is not statistically significant. If that coefficient were statistically significant, it could be stated that the fewer innovative lemmas a scribe used, the higher his spelling correctness tended to be. With regard to the category 0.02% to 0.40%, the pattern is as expected, whereas it remains unclear why the complete absence of (the seventy-nine) innovative lemmas does not lead to a statistically significant association with spelling correctness when compared to the baseline category. This may be related to document types, as innovative lemmas are more frequent in certain types of documents than they are in others. It might be beneficial to include the document type as a variable in a future study.

3) The greater the number of **Genitive plural forms**, the greater the level of correct spelling. The pattern is as expected: the genitive plural was likely to have been considered a conservative prestige form associated with high levels of correct spelling (the largest coefficient).

4) The greater the number of **Neuters in -*um* instead of -*o*/-*u*/-*us***, the higher the level of correct spelling. The pattern is as expected: the -*um* ending of the neuter was likely to have been considered a conservative prestige form associated with high levels of correct spelling.

5) The greater the number of ***De* PPs instead of genitives**, the lower the level of correct spelling. The pattern is as expected: the *de* PP was likely to have been considered an innovative low-prestige construction when compared to the genitive case. Consequently, *de* PPs predict low levels of correct spelling.

6) The greater the number of **Complementiser clauses instead of ACI**, the lower the level of correct spelling. The pattern is as expected: ACI was likely to have been considered a conservative prestige construction when compared to the complementiser clause. Consequently, complementiser clauses seem to predict low levels of correct spelling.

The first conclusion that can be drawn from the full model is that the statistical analysis partly confirms the working hypothesis presented in Section 1: good spelling is actually associated to some features that I defined as indications of the scribes' L2 competence in Section 4. This is shown by the meaningful and statistically significant associations between correct spelling and five of the linguistic variables that represent the features the scribes had to learn to use or to avoid in order to write in standard Latin. On the other hand, correct spelling

showed no statistically significant association with the other four features. The next step is to find an answer to the following questions: Do the statistically significant associations have something in common? Equally importantly, do the statistically non-significant associations have something in common?

A preliminary interpretation is that the features seem to be grouped according to their grammatical domain: standard Latin spelling appears to be associated in a meaningful and statistically significant way mainly with lexical (Innovative lexicon) and predominantly morphology-related features (Genitive plural, Neuter *-um*, *De* PP and Complementiser clauses), whereas predominantly syntactic features (Subjects/objects in *-bus*, Accusative subjects, OV order and Past participle/auxiliary order) have no statistically significant association with correct spelling. At this point, it must be emphasised that Latin uses inflections to mark the relationship between the word-level grammatical structure and the sentence-level structure. Therefore, it is sometimes difficult to draw a line between morphology and syntax (see the discussion on *-bus* in Section 6.2). In this study, 'morphological features' or 'morphology-related features' refer to those features that contain a functional morpheme but are not related to syntactic functions.

It appears that L2 syntax was not acquired as successfully as were the less complex L2 grammar domains of lexicon and morphology. Consequently, even good spellers used classical Latin syntactic features haphazardly. In the following Sections, I seek to show that the way in which scribal education dealt with language (Section 6.1) and the complexity of the grammatical domain (Section 6.2) are likely to have determined why the non-syntactic features were attributed a prestige status and thus appear to be associated with spelling correctness, which is considered to be an indicator of language competence in the present study.

6 Discussion

6.1 Scribal education

In Section 4, I proposed in passing that instruction in schools may have kept the neuter category alive artificially by reminding students of its existence and by making pupils memorise lists of neuters. Donatus' *Ars minor*, the basic pedagogical grammar of the time, defined neuter nouns by the sentence quoted in excerpt (2). Therefore, at least some future scribes may have learnt to consider the neuter *-um* ending to be a conservative feature. I (Korkiakangas 2010: 139) showed for the papyri of Ravenna (legal documents mainly from the sixth and

seventh centuries) and Löfstedt (1961: 226) for the juridical Latin of the *Edictum Rothari* (seventh century) that the neuter *-um* was largely maintained and more often than was the accusative *-um* of the masculine. This, and the fact that *-um* was also used hypercorrectly for the dative/ablative *-o* with neuters but not as often with masculines, supports my hypothesis.

> (2) Donatus, *Ars minor* (*GL* 4.356.10–11)
> scamnum nomen appellatiuum generis neutri numeri singularis figurae simplicis, casus nominatiui accusatiui et uocatiui
> *scamnum* [bench] is a neuter-gender singular-number non-compound common name, the form of the nominative, accusative and vocative case

Relatively little is known about early medieval Tuscan scribes' education (Mencacci 2012: 75–90; Petrucci 1973: 639; Schiaparelli 1924: 56ff.). An LLCT document from 767 CE (*ChLA* 997) mentions a school in the portico of the cathedral of Lucca. This school, which was obviously run by the church, seems to have provided clerics and some laymen with education in writing Latin for the purposes of ecclesiastical administration and/or documentary production (Coturri 1973: 695ff.). Initially, most documentary scribes were clerics. In the ninth century, when documentary production passed into lay hands (Keller 1973: 120–24), it is likely that the education of the scribes changed accordingly. This administrative reform seems to be reflected in the increasing uniformity of the documents, even in terms of improved spelling (Korkiakangas 2017). Schiaparelli (1924: 57) identified a few documentary scribes whose texts are included in the LLCT documents among the copyists who compiled the famous Manuscript 490, which is the only surviving book copied with certainty in Lucca during the eighth century. Thus, at least some Luccan scribes also knew book scripts, arguably a sign of higher literary education.

What the actual school curriculum in early medieval Tuscany included is unknown but, considering the tremendous linguistic variety in the LLCT documents, not much was required from all those who wrote documents. The grammatical treatises of the time can be used to approximate the maximum level that a researcher can expect from scribal education in early medieval Tuscan schools. The treatises that were available before the Carolingians began to promote Priscian's *Institutiones grammaticae* in the late eighth century only dealt with spelling, the parts of speech and the figures of speech (Black 2001: 64; Luhtala 1993: 145–6). The available grammatical treatises had been written in late antiquity when the distance between the syntax of written and spoken varieties of Latin was still

relatively small. Therefore, the treatises were concerned with syntactic issues only in passing, for example, when they discussed rhetorical figures.

Coturri (1973: 696) maintained that, in addition to the above-mentioned school, Lucca also had an elementary writing school, which would explain the high number of literate, but apparently not very experienced, people in Lucca, as evidenced by the large number of autograph subscriptions in the LLCT documents (cf. Petrucci 1973: 640). Technically, any higher literary education, such as the adoption of documentary formulae, must have been based on immersion in extant authoritative documents and on their subsequent imitation, either independently or in the classroom. This method had been in use since ancient times and seems to have worked without many difficulties in the culture of early medieval Europe (Black 2001: 64–70, cf. Cribiore 2001: 132–4). This was possible because the gap between the phonology and grammar of the spoken language and of classical Latin was still narrow enough not to compromise communication. Apparently, a superficial memorization of the classical spellings of frequent words and some understanding of parts of speech and their inflections occasionally passed for sufficient Latin literacy. Similarly, the imitation of extant texts was evidently sufficient to prevent the scribes from lapsing too often into non-classical syntactic choices, as shown in the relatively high percentages of the conservative variants of the syntactic variables in Section 4.

Although the learning method based on memorization and imitation still worked in early medieval Tuscany, it is likely that Latin L2 acquisition became increasingly challenging over the centuries, which widened the gap between the written standard Latin and the vernacular. By the early Middle Ages, an Italian Latin learner would have needed a basic understanding of syntactic functions to be able to cope with classical Latin word order and subject case marking, but school instruction did not foster such an understanding. The recognition of this syntactic gap may have been a stimulus that interested Carolingian scholars in Priscian's work at the end of the eighth century. Luhtala (1993: 161ff.) showed that the Carolingian grammarians elaborated on Priscian's (infant) syntactic analysis of subject and predicate. Black (2001), however, maintained that this scholarly interest had little or no effect on the everyday pedagogical grammar of Latin before the twelfth century, when it became necessary to adapt Latin teaching practices to the needs of larger audiences and when the advances in logic introduced the Aristotelian notions of subject and predicate into real grammatical praxis (Black 2001: 70).

Thus, the observations made in the statistical analysis in Section 5 confirm the picture that arises from the close reading of ancient and early medieval

grammatical treatises: the pedagogical grammar was concerned with spelling and morphology-related issues. There had not been a need for syntactic explanations in the centuries during which the grammatical tradition took shape and, when the need arose, it was not recognised for centuries. In the meanwhile, schoolroom instruction was limited to easily understandable features, depending on the teacher's skills.

On the other hand, it would be naïve to claim that the early medieval Latin educational system transmitted theoretical grammatical knowledge in such an efficient and uniform manner that it determined the scribes' writing performances in the way they are seen in the statistical analysis above. Studies of modern L2 acquisition show that, despite effective and theoretically grounded school instruction, acquisition is dependent on several other factors outside of the curriculum content, including students' motivation, learning skills, exposure time, the distance between the L1 and the L2 and, importantly, the properties of the linguistic features to be learnt (e.g. Lightbown and Spada 2006: 53ff., 82ff.; Matras 2009: 69–70; Jarvis and Odlin 2000). In fact, learning depends ultimately on the learner's personal *intake*, not the *input* provided, however adequate the latter may be. As stated above, early medieval L2 learners were supposed to immerse themselves in extant texts and to imitate them in order to reach the higher levels of literacy; in other words, recognised syntax and style.

It is also apparent that much of what was taught of morphology was left to be internalised by this immersion. No basic grammar guided L2 Latin learners to use the genitive case instead of the *de* PP in certain semantic contexts, or not to use a complementiser clause if one could use an ACI; that is, mainly with speech verbs. Grammatical treatises and, assumedly, teachers made pupils memorise lists of prepositions and complementisers (and probably neuters), but the understanding of the contexts in which these were to be used had to be adopted largely by immersion in texts. The following Section discusses how the perceptual properties of the linguistic features to be learnt may have affected the learning processes and, consequently, the patterning of the features examined in the present study.

6.2 Grammatical domains

As observed in Section 5, the statistically significant and non-significant associations of the full model seem to be grouped according to each feature's grammatical domain. I suggest here that this split between syntax and other grammatical domains is closely related not only to medieval scribes' education,

but also to the linguistic properties of the features themselves. Thus, teaching and linguistic properties are not to be considered mutually exclusive, but complementary, explanations: what was taught was likely to reflect the relative order of the complexity of linguistic domains, albeit unintentionally (see Table 2). Similarly, prescriptive instruction in school is bound to interfere with the linguistic motivation of the pattern observed in the statistical analysis.

Cognitive linguistics offers a useful framework for systematizing the split mentioned above. Cognitive theories do not assume clear-cut boundaries between what are traditionally called lexicon, morphology and syntax, but consider them to form a continuum (Broccias 2012). According to Croft and Cruse (2004: 247ff.), language is a repository of constructions of varying degrees of generality. This is illustrated by the syntax-lexicon continuum, a scale that locates language domains on a continuum according to their atomicity/complexity and specificity/schematicity (Table 2). Atomicity refers to the indivisibility into meaningful parts and is the opposite of complexity, while schematicity means the generality of category and is the opposite of specificity.

Table 2. Syntax-lexicon continuum (after Broccias 2012: 738 and Croft and Cruse 2004: 255).

Grammar domain	Traditional name	Example
Complex and (mostly) schematic	Syntax	noun verb noun (= transitive construction)
Complex and (mostly) specific	Idiom	*pull one's leg*
Complex but bound	Morphology	noun-*s*
Atomic and schematic	Word class	pronoun, adjective
Atomic and specific	Word/lexicon	*this*, *green*

The linguistic features were presented in Section 4 and examined in Section 5 in a tentative order of increasing grammatical complexity in terms of the syntax-lexicon continuum. The features belong to the domains Word/lexicon, Morphology and Syntax. As seen previously, the statistically significant associations are related to the domains Word/lexicon and Morphology, and the statistically non-significant associations to the complex and schematic domain of Syntax.

I argue elsewhere (Korkiakangas 2018: 433–4) that the order of the domains in the syntax-lexicon continuum reflects, among other things, the cognitive effort involved in recognizing those domains during L2 acquisition. The more complex domains (such as Syntax) are not recognised and, consequently, are not adopted as easily as are the less complex domains (such as Word/lexicon). However,

the evidence of the statistical analysis, as well as that gathered in Korkiakangas (2018), suggests that the complexity, as understood in terms of the syntax-lexicon continuum, is caused by various underlying motivations that can be identified and explained. This view implies that the complexity hierarchy inherent in the syntax-lexicon continuum should possibly be considered a useful but epiphenomenal simplification. In the following passages, I discuss one underlying motivation that is likely to be relevant for L2 acquisition of documentary Latin. This motivation is salience.

The complexity dimension of the syntax-lexicon continuum seems to be partly related to the perceptual properties of each domain's typical representatives (auditory properties in the oral context and visual properties in the written context). Lexical items and longer morphological inflections tend to be easily perceptible pieces of phonetic/graphic substance that a language learner recognises and adopts simply by memorization. This prominence or noticeability of linguistic units, such as words and (other) morphemes, is termed 'perceptual salience' (Cintrón-Valentín and Ellis 2016: 3; MacLeod 2015). Note that several other motivations, such as token frequency (e.g. Bybee 2007), cause linguistic units to be salient. However, the present research setting only applies to perceptual salience.

Contrary to lexicon and morphology, syntax pertains to the rules that operate behind the material façade of words and morphemes. The rules do not often have manifestations that are noticed easily, or the rationale for the manifestations is not easily deduced from the use contexts without specific linguistic intuition or without the help of an experienced teacher. A learner of Latin syntax must be capable of mapping information conveyed by the morphological inflections of words with the functions those words carry at the sentence level. This is much more difficult than is adopting individual non-contextualised words or morphemes. Consequently, syntactic features are at risk of passing unnoticed in superficial language acquisition. For example, an L2 learner of English learns lexical items and free functional morphemes, such as prepositions, quickly, but has to invest considerably more cognitive effort in learning to put *-s* at the end of those present tense verbs that are in the third person singular (cf. Zobl and Liceras 1994: 169–71).

The interpretation presented above is supported by several experimental studies on the L2 acquisition order of grammatical categories in modern languages, such as Zobl and Liceras (1994) and Goldschneider and DeKeyser (2001). Processability-informed theories assume that learners first acquire the relationships between lemmas, then those within words (lexical morphology),

within phrases, between phrases and, finally, between clauses (Pienemann 1999: 7–9). The studies by Goldschneider and DeKeyser (2001) and Cintrón-Valentín and Ellis (2016) confirm that perceptual salience plays an important role in L2 morpheme acquisition.

Free morphemes, such as words, the preposition *de* and complementisers, as well as the phonetically/graphically substantial genitive plural ending (e.g. *-orum*), are readily considered perceptually salient. It is also easy to agree that word order issues (OV order and Past participle/auxiliary verb order) are not determined by perceptually salient units, but by the linearization of syntactic functions. Even though Subject case encoding (Accusative subjects) does rely on case inflections, nominative and accusative, and the morphological difference between the nominative and accusative inflections of type *virtus – virtute*(*m*) may be perceptual as such, it is hard to decide which form is the salient one (perhaps *virtus*; cf. Korkiakangas 2018: 444).

There is, however, no need to speculate on the salience of the nominative and accusative forms, given that OV order, Past participle/auxiliary verb order and Subject case encoding are all based on syntactic rules, and syntactic rules require an understanding of abstract categories, such as syntactic function, which cannot be tackled in terms of salience. In order to assign the nominative case classically to all the subjects of finite verbs contrary to the late Latin morphosyntactic alignment, a late Latin speaker had to understand what the subject was. The knowledge of syntactic functions was also necessary in order to avoid the ablative inflection *-bus* with subjects and objects. The relatively salient ending *-bus* was possibly erroneously considered to be a prestige feature and was misused hypercorrectly in any syntactic context, even by good spellers.

However, not all the features examined in the statistical analysis seem to fit in with this explanation. The neuters in *-um* instead of *-o*/*-u*/*-us* can hardly be described as perceptually salient. The pronunciation of the ending is likely to have been /o/ while, in writing, <um> does not stand out due to its brevity, at least in comparison to the genitive plural ending, *-orum,* as four characters are considerably more than are two characters. Moreover, the neuter is not the only habitat of *-um*, which is also the classical accusative singular form of the *o*-stem masculine as well as the genitive plural form of the third declension. Thus, the perceptual properties of the *-um* ending are unlikely to have guaranteed it the position of being recognised as a prestige feature and the ensuing association with correct spelling. Nonetheless, the *-um* neuters are strongly associated with correct spelling; thus, the prestige attribution must derive from elsewhere, namely from prescriptive school instruction.

To summarise, both scribal education and linguistic properties are likely to have contributed to the diverse associations between correct spelling and the linguistic features observed in Section 5. On one hand, instruction in schools tends to sanction some linguistic features instead of others. On the other hand, an L2 learner naturally tends to pay more attention to easily perceivable, salient, linguistic features. Nor is it an accident that the split arises between syntax and other domains for both criteria. It is only to be expected that the concerns of the grammatical tradition are primarily and originally related to the perceptually most salient, that is, the most easily manageable, domains, whereas syntax receives attention only as a result of the theoretical development of the field.

7 Conclusion

This study examined whether the variations in spelling and the variations in the use of certain linguistic features were associated with each other in early medieval documentary Latin as documented in LLCT. This was accomplished by quantifying spelling variations and selected linguistic features that are known to be indicative of prestige (conservative, classical Latin features) or the lack thereof (innovative, Romance-type features) and by the subsequent analysis of their statistical associations using a general linear model.

The statistical analysis confirmed the working hypothesis that correct spelling and learned grammatical competence are associated. However, statistically significant associations are limited to lexical and morphology-related features. This suggests that L2 syntax was not acquired as successfully as were the less complex L2 grammar domains. I explained this pattern via the complexity of each grammatical domain in question and via the prestige assessments derived from prescriptive school instruction. The pedagogical grammatical tradition available at the time was not concerned with syntax, but concentrated on lexicon, spelling and morphology instead. The correct contexts of use were expected to be learnt by imitating authoritative models, and it is here that the grammatical domains seem to play a role.

The syntax-lexicon continuum, a framework adopted from cognitive linguistics, ranks different grammatical domains according to their complexity, which seems to determine the ease with which L2 learners acquire them. I have suggested, however, that the fundamental reason behind this ranking, as far as L2 acquisition is concerned, might be the fact that the domains tend to be represented differently in perceptual terms: perceptually salient lexical items and

morphological morphemes are easy to notice and acquire, while syntax is much more complex in the sense that it requires an understanding of the relationship between words and their inflections. Consequently, an L2 learner succeeded in associating prestige (such as with the genitive plural form) or the absence thereof (such as with the *de* PP) with perceptual features. For syntactic features, the prestige/no prestige attribution often failed, hence the statistically non-significant associations with correct spelling. The salience-based explanation obtains support from several studies on the L2 acquisition order of grammatical categories in modern languages. To achieve a more detailed picture of the role of salience, a further study of the association between correct spelling and lowly salient morphological features other than *-um* would be required. In fact, Korkiakangas (2018) suggested that morphological features that are particularly non-salient easily escaped the scribes' attention.

Thus, school instruction and linguistic properties are not to be considered mutually exclusive explanations, but are both likely to contribute to the differing associations between spelling and linguistic features in the LLCT. Although the exact relationship between domain complexity and salience cannot be fully established within the present research setting, they both seem to have an important role in how the LLCT scribes acquired their Latin. In summary, spelling correctness can duly be utilised as a rough measure of early medieval scribes' Latin competence.

Online resources

LLCT = Late Latin Charter Treebank (LLCT1, version 2) is available at https://zenodo.org/record/3633607#.YlUMr4VBxPY.

Open Office Latin lexicon is available at https://github.com/cisocrgroup/Resources/tree/master/lexica/latin.

Whitaker's WORDS is available at http://mk270.github.io/whitakers-words/operational.html.

References

Adams, J. N. 2013. *Social Variation and the Latin Language*. Cambridge University Press.

Amsler, Mark E. 1993. 'History of Linguistics, "Standard Latin", and Pedagogy', in V. A. Law (ed.), pp. 49–66.

Bamman, David, Passarotti, Marco, Crane, Gregory and Raynaud, Savina 2007. *Guidelines for the Syntactic Annotation of Latin Treebanks*, v. 1.3. (https://static.perseus.tufts.edu/docs/guidelines.pdf, accessed 12 April 2022)

Bartoli Langeli, Attilio 2006. *Notai: scrivere documenti nell'Italia medievale*. Rome: Viella.

Bauer, Brigitte L. M. 2006. '"Synthetic" vs. "Analytic" in Romance: the Importance of Varieties', in R. Gess and D. Arteaga (eds.), *Historical Romance Linguistics: Retrospective and Perspectives*, pp. 287–304. Amsterdam: John Benjamins.

Black, Robert 2001. *Humanism and Education in Medieval and Renaissance Italy*. Cambridge University Press.

Broccias, Cristiano 2012. 'The Syntax-Lexicon Continuum', in Terttu Nevalainen and Elizabeth Closs Traugott (eds.), *The Oxford Handbook of the History of English*, pp. 735–47. Oxford University Press.

Burt, Jennifer S. 2006. 'Spelling in Adults: The Combined Influences of Language Skills and Reading Experience', *Journal of Psycholinguistic Research* 35: 447–70.

Bybee, Joan 2007. *Frequency of Use and the Organization of Language*. Oxford University Press.

Carey, Gregory 2013. *Quantitative Methods in Neuroscience*. Online book draft. (http://psych.colorado.edu/~carey/qmin/QMIN_2013_03_17.pdf, accessed 12 April 2022)

Castellani, Arrigo 1976. *I più antichi testi italiani*. Bologna: Pàtron.

Cennamo, Michela 2009. 'Argument Structure and Alignment Variations and Changes in Late Latin', in Jóhanna Barðdal and Shobhana L. Chelliah (eds.), *The Role of Semantic, Pragmatic, and Discourse Factors in the Development of Case*, pp. 307–46. Amsterdam: John Benjamins.

Cintrón-Valentín, Myrna C. and Ellis, Nick C. 2016. 'Salience in Second Language Acquisition: Physical Form, Learner Attention, and Instructional Focus', *Frontiers in Psychology* 7: 1284.

Coturri, Enrico. 1973. 'Fonti per uno studio della cultura a Lucca dall'VIII all'XI secolo', in *Atti del V Congresso internazionale di studi sull'alto medioevo*, pp. 695–702. Spoleto: Centro Italiano di Studi sull'Alto Medioevo.

Cribiore, Raffaella 2001. *Gymnastics of the Mind: Greek Education in Hellenistic and Roman Egypt*. Princeton University Press.

Croft, William and Cruse, D. Alan 2004. *Cognitive Linguistics*. Cambridge University Press.

Danckaert, Lieven 2016. 'Variation and Change in Latin BE-Periphrases: Empirical and Methodological Considerations', in J. N. Adams and Nigel Vincent (eds.), *Early and Late Latin: Continuity or Change?*, pp. 132–62. Cambridge University Press.

de Dardel, Robert and Gaeng, Paul A. 1992. 'La déclinaison nominale du latin non classique: essai d'une méthode de synthèse', *Probus* 4: 91–125.

de Paolis, Paolo 2010. 'L'insegnamento dell'ortografia latina fra Tardoantico e alto Medioevo: teorie e manuali', in Lucio Del Corso and Oronzo Pecere (eds.), *Libri di scuola e pratiche didattiche dall'antichità al rinascimento*, pp. 229–91. Università degli studi di Cassino.

Fiorentino, Giuliana 1994. 'Aspetti della morfosintassi nominale nelle carte cavensi del IX secolo', *Archivio Glottologico Italiano* 79: 23–50.

GL = *Grammatici Latini*, ed. Heinrich Keil, Leipzig: Teubner 1855–80.

Goldschneider, Jennifer M. and DeKeyser, Robert M. 2001. 'Explaining the "Natural Order of L2 Morpheme Acquisition" in English: A Meta-analysis of Multiple Determinants', *Language Learning* 51: 1–50.

Greco, Paolo 2017. 'La complementazione frasale', in Rosanna Sornicola, Elisa D'Argenio and Paolo Greco (eds.), *Sistemi, norme, scritture: la lingua delle più antiche carte cavensi*, pp. 267–310. Naples: Giannini.

Herman, József 1991. 'Spoken and Written Latin in the Last Centuries of the Roman Empire: A Contribution to the Linguistic History of the Western Provinces', in Wright (ed.), pp. 29–43.

Herman, József 1992. 'Sur quelques aspects du latin mérovingien: langue écrite et langue parlée', in Maria Iliescu (ed.), *Latin vulgaire – latin tardif III: Actes*

du IIIème Colloque international sur le latin vulgaire et tardif, pp. 173–86. Tübingen: Max Niemeyer.

Jarvis, Scott and Odlin, Terence 2000. 'Morphological type, spatial reference, and language transfer', *Studies in Second Language Acquisition* 22: 535–55.

Keller, Hagen 1973. 'La marca di Tuscia fino all'anno Mille', in *Atti del V Congresso internazionale di studi sull'alto medioevo*, 117–36. Spoleto: Centro Italiano di Studi sull'Alto Medioevo.

Korkiakangas, Timo 2010. 'Neutro plurale e femminile singolare: il fattore grafico nell'interpretazione delle peculiarità della declinazione tardolatina', *Arctos* 44: 133–45.

Korkiakangas, Timo 2016. *Subject Case in the Latin of Tuscan Charters of the 8th and 9th Centuries*. Helsinki: Societas Scientiarum Fennica.

Korkiakangas, Timo 2017. 'Spelling Variation in Historical Text Corpora: The Case of Early Medieval Documentary Latin', *Digital Scholarship in the Humanities* 33: 575–91.

Korkiakangas, Timo 2018. 'Spoken Latin Behind Written Texts: Formulaicity and Salience in Medieval Documentary Texts', *Diachronica* 35: 429–49.

Korkiakangas, Timo and Lassila, Matti 2013. 'Abbreviations, Fragmentary Words, Formulaic Language: Treebanking Medieval Charter Material', in Francesco Mambrini, Marco Passarotti and Caroline Sporleder (eds.), *Proceedings of the Third Workshop on Annotation of Corpora for Research in the Humanities*, pp. 61–72. Sofia: Bulgarian Academy of Sciences.

Korkiakangas, Timo and Passarotti, Marco 2011. 'Challenges in Annotating Medieval Latin Charters', *Journal of Language Technology and Computational Linguistics* 26: 103–14.

Larson, Pär 2012. 'Le carte alto-medievali come fonte di lingua: qualche esperienza personale', in Rosanna Sornicola and Paolo Greco (eds.), *La lingua dei documenti notarili alto-medievali dell'Italia meridionale: bilancio degli studi e prospettive di ricerca*, pp. 63–73. Naples: Tavolario.

Law, Vivien A. 1982. *The Insular Latin Grammarians*. Woolbridge: The Boydell Press.

Law, Vivien A. (ed.) 1993. *History of Linguistic Thought in the Early Middle Ages*. Amsterdam – Philadelphia: John Benjamins.

Ledgeway, Adam 2012. *From Latin to Romance: Morphosyntactic Typology and Change*. Oxford University Press.

Lightbown, Patsy M. and Spada, Nina 2006. *How Languages are Learned*. 2nd edition. Oxford University Press.

Löfstedt, Bengt 1961. *Studien über die Sprache der langobardischen Gesetze*. Stockholm – Göteborg – Uppsala: Almqvist & Wiksell.

Lüdtke, Helmut 1964. 'Die Entstehung romanischer Schriftsprachen', *Vox Romanica* 23: 3–21.

Luhtala, Anneli 1993. 'Syntax and Dialectic in Carolingian Commentaries on Priscian's *Institutiones grammaticae*', in V. A. Law (ed.), pp. 145–91.

MacLeod, Bethany 2015. 'A Critical Evaluation of Two Approaches to Defining Perceptual Salience', *Ampersand* 2: 83–92.

Matras, Yaron 2009. *Language Contact*. Cambridge University Press.

Mencacci, Paolo 2012. *Vita civile ed ecclesiastica a Lucca nell'alto Medioevo, secoli VI–IX*. Lucca: Pacini Fazzi.

Petrucci, Armando 1973. 'Scrittura e libro nella Tuscia altomedievale (secoli VIII–IX)', in *Atti del V Congresso internazionale di studi sull'alto medioevo*, pp. 627–43. Spoleto: Centro Italiano di Studi sull'Alto Medioevo.

Pienemann, Manfred 1999. *Language Processing and Second Language Development: Processability Theory*. Amsterdam: John Benjamins.

Politzer, Frieda N. and Politzer, Robert L. 1953. *Romance Trends in 7th and 8th Century Latin Documents*. Chapel Hill: The University of North Carolina Press.

Rovai, Francesco 2012. *Sistemi di codifica argomentale: tipologia ed evoluzione*. Pisa: Pacini.

Sabatini, Francesco 1965. 'Esigenze di realismo e dislocazione morfologica in testi preromanzi', *Rivista di Cultura Classica e Medievale* 7: 972–98.

Sairio, Anni and Palander-Collin, Minna 2012. 'The Reconstruction of Prestige Patterns in Language History', in Juan Manuel Hernández-Campoy and Juan Camilo Conde-Silvestre (eds.), *The Handbook of Historical Sociolinguistics*, pp. 626–38. Chichester: Blackwell.

Schiaparelli, Luigi 1924. *Il codice 490 della Biblioteca capitolare di Lucca*. Pompeo Sansaini Editore: Rome.

Schiaparelli, Luigi 1933. 'Note diplomatiche sulle carte longobarde II: Tracce di antichi formulari nelle carte longobarde', *Archivio storico italiano* 19: 3–34.

Sornicola, Rosanna 2012. 'Bilinguismo e diglossia dei territori bizantini e longobardi del Mezzogiorno: le testimonianze dei documenti del IX e X secolo', *Quaderni dell'Accademia Pontaniana* 59: 1–102.

Spevak, Olga 2010. *Constituent Order in Classical Latin Prose*. Amsterdam: John Benjamins.

Valentini, Cecilia 2018. *L'evoluzione della codifica del genitivo dal tipo sintetico al tipo analitico nelle carte del* Codice diplomatico longobardo. Università degli Studi di Firenze.

Walsh, Thomas J. 1991. 'Spelling Lapses in Early Medieval Latin Documents and the Reconstruction of Primitive Romance Phonology', in Wright (ed.), pp. 205–18.

Wright, Roger (ed.) 1991. *Latin and the Romance Languages in the Early Middle Ages*. The Pennsylvania State University Press.

Wright, Roger 1991. 'The Conceptual Distinction Between Latin and Romance: Invention or Evolution', in Wright (ed.), pp. 103–13.

Zamboni, Alberto 2000. *Alle origini dell'italiano: dinamiche e tipologie della transizione dal latino*. Rome: Carocci.

Zobl, Helmut and Liceras, Juana 1994. 'Functional Categories and Acquisition Orders', *Language Learning* 44: 169–80.

Index

Subject Index

Commentationes Humanarum Litterarum

Vol. 90 (1990)

Solin, Heikki: *Namenpaare. Eine Studie zur römischen Namengebung*. 1990. 92 p.

Vol. 91 (1990)

Roman Eastern Policy and Other Studies in Roman History, edited by Heikki Solin and Mika Kajava. 1990. 174 p.

Vol. 93 (1991)

Bruun, Christer: *The Water Supply of Ancient Rome. A Study of Roman Imperial Administration*. 1991. 468 p.

Vol. 94 (1991)

Koskenniemi, Erkki: *Der philostrateische Apollonios*. 1991. 101 p.

Vol. 96 (1991)

Forsén, Björn: *Lex Licinia Sextia de modo agrorum – fiction or reality?* 1991. 88 p.

Vol. 97 (1992)

Salomies, Olli: *Adoptive and Polyonymous Nomenclature in the Roman Empire*. 1992. 179 p.

Vol. 102 (1994)

Leiwo, Martti: *Neapolitana. A Study of Population and Language in Graeco-Roman Naples*. 1994. 232 p.

Vol. 103 (1995)

Chydenius, Johan: *The Spirituality of Fénelon in his Latin Writings 1700–1712*. 1995. 60 p.

Vol. 104 (1995)

Acta colloquii epigraphici Latini Helsingiae 3.–6. sept. 1991 habiti, ediderunt Heikki Solin, Olli Salomies, Uta-Maria Liertz. 1995. 425 p.

Vol. 108 (1996)

Mehtonen, Päivi: *Old Concepts and New Poetics. Historia, Argumentum, and Fabula in the Twelfth- and Early Thirteenth-Century Latin Poetics of Fiction*. 1996. 173 p.

Vol. 109 (1996)

Kuisma, Oiva: *Proclus' Defence of Homer*. 1996. 157 p.

Vol. 111 (1998)

Lampela, Anssi: *Rome and the Ptolemies of Egypt. The Development of their Political Relations 273–80 B.C.* 1998. 301 p.

Vol. 112 (1998)

Nummenmaa, Tapio: *Divine Motions and Human Emotions in the Philebus and in the Laws. Plato's Theory of Physic Powers*. 1998. 151 p.

Vol. 113 (1999)

Thesleff, Holger: *Studies in Plato's Two-Level Model*. 1999. 143 p.

Vol. 114 (1999)

Dimitropoulos, Panagiotis: *Untersuchungen zum finalen Genetiv des substantivierten Infinitivs bei Thukydides*. 1999. 117 p.

Vol. 115 (2000)

Heinonen, Sirkka: *Prometheus Revisited. Human Interaction with Nature through Technology in Seneca*. 2000. 232 p.

Vol. 116 (2000)

Lehtonen, Tuomas M. S. & Mehtonen, Päivi (eds.): *Historia. The Concept and Genres in the Middle Ages*. 2000. 142 p.

Vol. 118 (2002)

Kivistö, Sari: *Creating Anti-eloquence. Epistolae obscurorum virorum and the Humanist Polemics on Style*. 2002. 256 p.

Vol. 119 (2002)

Salmenkivi, Erja: *Cartonnage Papyri in Context. New Ptolemaic Documents from Abū Sir al Malaq*. 2002. 182 p.

Vol. 120 (2003)

Kuisma, Oiva: *Art or Experience. A Study on Plotinus' Aesthetics*. 2003. 207 p.

Vol. 121 (2004)

Korhonen, Kalle: *Le iscrizioni del Museo civico di Catania*. 2004. 418 p.

Vol. 122: 1–2 (2007)

Proceedings of the 24th International Congress of Papyrology. Helsinki, 1–7 August, 2004, edited by Jaakko Frösén, Tiina Purola and Erja Salmenkivi. 2007. 1075 p.

Vol. 123 (2007)

Steinby, Christa: *The Roman Republican Navy*. 2007. 236 p.

Vol. 124 (2009)

Halla-aho, Hilla: *The non-literary Latin letters. A study of their syntax and pragmatics*. 2009. 190 p.

Vol. 125 (2008)

The Konikovo Gospel (Bibl. Patr. Alex. 268), edited by Jouko Lindstedt, Ljudmil Spasov and Juhani Nuorluoto. 2008. 439 p.

Vol. 126 (2009)

Sipilä, Joonas: *The Reorganisation of Provincial Territories in Light of the Imperial Decision-making Process. Later Roman Arabia and Tres Palaestinae as Case Studies*. 2009. 328 p.

Vol. 127 (2010)

Saastamoinen, Ari: *The Phraseology of Latin Building Inscriptions in Roman North-Africa*. 2010. 646 p.

Vol. 128 (2011)

Building Roma Aeterna. Current Research on Roman Mortar and Concrete. Proceedings of the conference, March 27–29, 2008. Editors Åsa Ringbom and Robert L. Hohlfelder, assistant editors Pia Sjöberg and Pia Sonck-Koota. 2011. 260 p.

Vol. 129 (2011)

Eikonopoiia. Digital Imaging of Ancient Textual Heritage. Proceedings of the international conference, Helsinki, 28–29 November, 2010, edited by Vesa Vahtikari, Mika Hakkarainen, Antti Nurminen. 2011. 267 p.

Vol. 130 (2013)

Maurizi, Luca: *Il cursus honorum senatorio da Augusto a Traiano. Sviluppi formali e stilistici nell'epigrafia latina e greca*. 2013. 324 p.

Vol. 131 (2014)

Papyri Turkuenses (P. Turku). Die Papyri im Besitz der Universitätsbibliothek Turku, herausgegeben von Heikki Koskenniemi unter Mitwirkung von Erkki Koskenniemi und Johannes Koskenniemi. 2014. 135 p.

Vol. 132 (2015)

Second Sailing: Alternative Perspectives on Plato. Edited by Debra Nails and Harold Tarrant in Collaboration with Mika Kajava and Eero Salmenkivi. 2015. 366 p.

Vol. 133 (2016)

Korkiakangas, Timo: *Subject Case in the Latin of Tuscan Charters of the 8th and 9th Centuries*. 2016. 276 p.

Vol. 134 (2018)

Kivistö, Sari: *Lucubrationes Neolatinae. Readings of Neo-Latin Dissertations and Satires*. 2018. 244 p.

Vol. 135 (2018)

Buchholz, Matias: *Römisches Recht auf Griechisch. Prolegomena zu einer linguistischen Untersuchung der Zusammensetzung und Semantik des byzantinischen prozessrechtlichen Wortschatzes*. 2018. 236 p.

Vol. 136 (2019)

Domus Pompeiana M. Lucretii IX 3, 5.24. The Inscriptions, Works of Art and Finds from the Old and New Excavations. Edited by Ria Berg and Ilkka Kuivalainen. 2019. 323 p.

Vol. 137 (2019)

Studi storico-epigrafici sul Lazio antico II. A cura di Heikki Solin. 2019. 168 p.

Vol. 138 (2020)

Meilicha Dôra. Poems and Prose in Greek from Renaissance and Early Modern Europe. Edited by Mika Kajava, Tua Korhonen and Jamie Vesterinen. 2020. 356 p.

Vol. 139 (2020)

Papers on Ancient Greek Linguistics. Proceedings of the Ninth International Colloquium on Ancient Greek Linguistics (ICACL 9), 30 August – 1 September 2018, Helsinki. Edited by Martti Leiwo, Marja Vierros & Sonja Dahlgren. 2020. 578 p.

Vol. 140 (2021)

Kuivalainen, Ilkka: *The Portrayal of Pompeian Bacchus*. 2021. 285 p.

Vol. 141 (2021)

Solin, Heikki: *Da Rodolfo Pio ai Farnese. Storia di due collezioni epigrafiche*. 2021. 432 p.

Vol. 142 (2022)

Kivistö, Sari: *Neo-Latin Verse Satire, ca. 1500–1800. An Ethical Approach*. 2022. 293 p.

Vol. 143 (2022)

Korhonen, Tua: *To the Glory that was Greece. Ideas, Ideals and Practices in Composing Humanist Greek during the Seventeenth Century*. 2022. 411 p.

Vol. 144 (2022)

Kajava, Mika: *Naming Gods. An Onomastic Study of Divine Epithets Derived from Roman Anthroponyms*. 2022. 159 p.

Vol. 145 (2023)

De Martino, Gianluca: *The Cult of Poseidoniate Hera and the Lucanians in Poseidonia/Paistom. An Ancient Story of Religion and Multiculturalism*. 2023. 329 p.

Vol. 146 (2023)

Nuorluoto, Tuomo. *Latin Female Cognomina. A Study on the Personal Names of Roman Women*. 2023. 529 p.