CHAPTER 7

Diminutives in the Greek papyri

A corpus-based investigation

Alek Keersmaekers
University of Leuven

Abstract

Diminutives are ubiquitous in papyri, but so far have been understudied. One reason for this is that it is rather difficult to retrieve a large and representative sample of diminutives, due to a lack of derivational annotation in papyrus corpora and the difficulty of separating non-diminutive and diminutive meanings of suffixes such as -ιον. One of the main aims of this paper is to describe a methodology to retrieve such a sample, making use of a large annotated corpus and techniques from distributional semantics in order to distinguish various meanings of diminutive suffixes. This method was mostly successful, although the small size and the formulaic nature of the papyrus corpus provide difficulties for distributional semantics. Additionally, this paper provides a first exploratory analysis of the factors conditioning variation between diminutive and non-diminutive forms from the framework of alternation studies, through the use of conditional inference trees. This analysis suggests that register, regional, and diachronic factors play a

How to cite this book chapter:
Keersmaekers, Alek. 2025. 'Diminutives in the Greek papyri: A corpus-based investigation'. In *Rays of Language: Linguistic Perspectives on Non-Literary Papyri and Related Sources*, edited by Klaas Bentein and Marja Vierros, 139–167. Helsinki: Helsinki University Press. https://doi.org/10.33134/HUP-34-7

role in this choice, with a major increase in the use of diminutives happening in the Fayum region, while cross-regionally diminutives are mostly confined to low-register texts.

7.1 Introduction

Fifteen years ago *The Language of the Papyri* was published, in which the editors state that '[w]e are still dealing today with linguistic resources of extraordinary richness which have hardly begun to be explored' (Evans and Obbink 2010, 2). Currently the linguistic analysis of Greek papyri is a booming field, and thanks to the papyri our knowledge about everyday Greek, sociolinguistics, and language change in the postclassical period has expanded greatly. The focus of this chapter is on diminutive suffixes, which, despite these developments, have received little attention so far.

Anyone who has intensively worked with papyri knows that diminutive suffixes are highly frequent in these texts: nouns ending in -ιον, for example, occur more than three times as often in papyri than in literary texts.[1] Such a highly frequent phenomenon lends itself perfectly to a corpus-based analysis. However, one of the reasons that this has not been attempted before is that it is quite difficult to retrieve diminutives. For example, a simple search for words ending in -ιον in the Papyrological Navigator at Papyri.info will retrieve many spurious hits such as ἄξιον, προσκεφάλαιον, and ἐπιτήδειον (and if other inflectional endings are added, the amount of false positives increases). Searching in lemmatised and morphologically annotated corpora, such as the duke-nlp corpus (also known as Trismegistos Words; Keersmaekers and Depauw 2022) already makes this easier, by limiting the search to nouns of which the lemma ends in -ιον or -ίον. This also returns numerous false positives, such as forms of δηνάριον and ἔλαιον. In other words, since the papyrus corpora lack derivational morphological annotation, a dedicated search strategy is necessary to retrieve diminutive forms.

This chapter therefore does not aim to offer a comprehensive study of diminutives in the Greek papyri or very definite results, but focuses primarily on the methodology to retrieve a large set of diminutives from the corpus. This is the subject of Section 7.3, after briefly discussing the literature on diminutives and how to define them in Section 7.2. Next, a preliminary analysis of this large dataset, using statistical

techniques, is given in Section 7.4. Section 7.5 further reflects on the main results, discusses possible problems with the methodology, and raises questions that are still to be answered.

7.2 Background: diminutives in Greek

As for the papyri, the only dedicated study on diminutive suffixes so far are the sections in Palmer's grammar of the post-Ptolemaic papyri (1945, 84–91), which mostly focuses on providing exhaustive lists of examples. For classical Greek, meanwhile, there are the extensive monographs of Petersen (1910, 1913); Swanson (1958) and Watt (2013) discuss diminutives in the New Testament and Meluzzi (2017) in Aristophanes; there is also an overview article by Grandi (2014), while Chantraine (1933) and Lühr and Balles (2008), offering a general treatment of derivation in Ancient Greek, also have some sections on diminutive suffixes.

One central issue is how to define a diminutive, given that on the one hand there are multiple suffixes in Greek that have diminutive meaning (e.g. -ιον, -ισκος, -αριον) and on the other hand typical 'diminutive' suffixes such as -ιον have a wide variety of meanings, not all of which are related to 'diminutive' semantics. It is generally accepted that the diminutive suffix -ιον derives from an adjectival function 'relating to' (Chantraine 1933, 64; Lühr and Balles 2008, 240; Petersen 1910, 132), which gave rise to a wide variety of meanings, including diminutives, but also words describing entirely different relationships with their base (e.g. ἴσθμιον 'necklace' from ἰσθμός 'neck'; Lühr and Balles 2008, 223) as well as locative nouns (e.g. χαλκεῖον 'forge' from χαλκεύς 'smith'; Lühr and Balles 2008, 222). As a consequence, the term 'diminutive' has been employed very broadly, referring to all words possessing the suffix -ιον, as well as more narrowly, referring to those forms that have an evaluative meaning (see Grandi 2014), that is, where a diminutive suffix is used to either comment on the physical properties of an entity ('small') or the speaker's attitude towards the entity ('good'/'bad') (Grandi 2014, 489). However, it is not always easy to draw a line between evaluative and nonevaluative diminutives (Grandi 2014, 489; Petersen 1910, 138–39).

This chapter focuses on variation between diminutives and their base forms (e.g. παιδίον 'little child' vs παῖς 'child'), from the field of alternation studies (e.g. Gries 2017; Pijpops 2020). This field, having its

roots in variationist sociolinguistics, investigates which factors condition the choice between variant constructions. If we employ this perspective, the assumption should be that diminutives and base forms are interchangeable to some extent, that is, they are variant constructions. Even if we limit ourselves to diminutives having evaluative meaning, the question whether the diminutive is interchangeable with its base form depends on the form under discussion, which I divide into three categories.

For some words (type A), the diminutive refers to an entity that is referentially distinct from the base form: Petersen (1910, 165), for example, cites the example of θυλάκιον 'seed capsule', that is a distinct concept from θύλακος 'bag', and the two words cannot be called interchangeable in any way (since θύλακος cannot be used to refer to the concept 'seed capsule').

In cases where the diminutive simply adds a meaning nuance (type B: the evaluative meanings discussed by Grandi 2014, e.g. 'little', 'dear'), it is debatable whether it is interchangeable with its base form. On the one hand, if we assume that the meaning is altered by using the diminutive suffix, it is referentially distinct from its base form and it cannot be called interchangeable, as the speaker has no choice, given that they cannot use the base form to express the same meaning they wanted to express. On the other hand, the nuances of meaning the diminutive suffix adds are related to the subjective perspective of the speaker (see e.g. Petersen 1910, 155) and are often employed for pragmatic effects; for example, the diminutive of an object being asked for is used to soften the request (cf. Petersen 1910, 163; Meluzzi 2017, 138–40). In such cases the diminutive could be called interchangeable to the extent that the base form could also be used without the meaning being substantially altered.

Finally, sometimes (type C) the diminutive suffix does not even add a meaning nuance, but the diminutive and the base are entirely interchangeable, such as πωλίον vs πῶλος (both 'foal') (Petersen 1910, 166).[2] In this case the diminutive obviously qualifies as a variant construction. It is important to point out that many words can belong to multiple classes: for example, the word θυλάκιον does not necessarily have to mean 'seed capsule' but is also used with the meaning 'little bag' (or simply 'bag'), as in P.Cair. Zen. 1 59069 (TM 725), 'θυλάκιον ἐσφραγισμένον α', for example.

For the papyri, it has been remarked that type C diminutives are rather frequent, although there are still contexts where it is clear that we are dealing with a type B diminutive (Amundsen 1965, 14–16). If the diminutive and base word are semantically equivalent, other factors than semantics might play a role: previous work has pointed out the role of extralinguistic factors such as register (e.g. Watt 2013, 67) or gender (e.g. Sommerstein 1995, although Meluzzi 2017 offers counterevidence). Such factors are the focus of this chapter, under the assumption that they are an important driving force for the distribution of diminutives in the papyri.

The analysis is carried out on a dataset containing both type B and type C diminutives, which are not distinguished; the procedure of extracting diminutives from the papyrus corpus and isolating type A diminutives is discussed in Section 7.3. The reason that this study makes no distinction between type B and C diminutives is that it is difficult to distinguish them both on a practical and on a conceptual level. On a practical level, there are no easy automated methods to do so (although it is possible for type A diminutives; see Subsection 7.3.2), and even when trying to distinguish them in a manual way, in most cases there are no overt contextual clues suggesting whether the diminutive was used to emphasise, for example, smallness or endearment or not. On a conceptual level, it is difficult to decide where to draw the boundary between diminutives that are semantically equivalent with their base word and those that are not, as was discussed earlier in this section.

This has important consequences from a statistical as well as a theoretical perspective. From a statistical perspective, it could be the case that since I did not exclude type B diminutives, patterns that seem to be related to extralinguistic factors might be related to semantic factors instead (e.g. texts in register X contain many diminutives because in register X small entities are more common, rather than because there is a real register difference). These problems are discussed in Section 7.5. From a theoretical perspective, this study conflates forms where there is a choice between diminutive suffix and base form (so, alternations) with those where there might not strictly be a choice. It should be pointed out, however, that for the field of alternation studies in general it has been difficult to define in which cases the language user has a choice (see Pijpops 2020, 286–89). This study therefore avoids taking an explicit stance on which diminutives are interchangeable and which

are not, while taking into account that the prevalence of diminutives in certain contexts does not always mean that the base form was an option that the language users could have chosen.

7.3 Methodology

7.3.1 Identifying possible candidates

To compile a list of (possible) diminutive suffixes, I relied on the suffixes mentioned as 'Diminutiva' in Lühr and Balles (2008, 240–46) (under the assumption that this work is exhaustive).[3] These are -ιον, -ισκος/-ισκη, -ιδιον, -αριον, -αδιον, -ακιον, -φιον, -ισκαριον, -ισκιον, -υδιον, -υδριον, -υλλιον, and -ις. I extracted all lemmas that end with one of these suffixes (controlling for accent variation, e.g. the search included both lemmas ending on -ιον and -ίον) from the duke-nlp automatically annotated corpus (Keersmaekers and Depauw 2022). Next, to control the size of this dataset and be able to assess the semantic value of the diminutive suffix (see Subsection 7.3.2), I filtered out all lemmas that occur fewer than ten times in the whole corpus. This yielded 333 unique lemmas (type frequency) of 69,897 tokens (token frequency). Not all of these lemmas were diminutives. I therefore manually checked the list and further reduced it on the basis of the following criteria:

- I removed coincidental matches. For example, one of the matches for -ισκος was δίσκος, which obviously does not have this suffix.
- I removed loanwords: typically the suffix is not native to Greek but from the source language (e.g. πραισίδιον 'garrison' from Latin *praesidium*). In some cases -ιον is also used as a suffix to create loans (see Palmer 1945, 86): Greek τούρτιον 'flatbread', for example, derives from Latin *torta*. It is doubtful whether -ιον has diminutive force here, but at any rate these examples were also excluded, since there was no Greek base without suffix to which these instances of -ιον could be compared.
- I removed words that had a deverbal (e.g. ἐντάγιον 'order for payment', from ἐντάσσω) or deadjectival (e.g. ἐγκύκλιον 'tax on sales', from ἐγκύκλιος) origin.
- I removed compounds when there was no Greek base without suffix. For example, although ἡμικάδιον ('half a kados', viz. a liquid

measure) could be called a diminutive, *ἡμίκαδος does not exist, but it is rather derived from ἡμι- + κάδιον. In some cases there is not even a diminutive base, for example, ἡμιαρτάβιον ('half an artaba', viz. a capacity measure) could similarly be analysed as ἡμι- + ἀρτάβιον, but *ἀρτάβιον does not occur in Greek. This might point to the use of -ιον as a compounding (rather than diminutive) suffix (or alternatively, ἡμιαρτάβιον might simply have been created on the basis of analogy with words such as ἡμικάδιον).

- I excluded instances of the suffixes -τηριον and -ειον (for places and instruments) such as πορθμεῖον 'landing place' and δεσμωτήριον 'prison' (Lühr and Balles 2008, 222). More doubtful were instances of -ιον that have similar functions, such as βιβλιοφυλάκιον 'place to keep books' and γεώργιον 'field' (cf. Lühr and Balles 2008, 222). Obviously, it is important to note that <ι> and <ει> are phonetically identical in postclassical Greek, so there is only a spelling difference between -ιον and -ειον, and the two suffixes often seem to be used interchangeably in papyri (see also Palmer 1945, 79). Nevertheless, I initially included these forms but marked them as 'locative -ιον suffix'.

This yielded the results presented in Table 7.1. As can be judged from this table, -ιον, -αριον, and -ιδιον are the main suffixes used to form diminutives in the Greek papyri.[4]

Table 7.1: Diminutive suffixes in the Greek papyri, for lemmas occurring at least ten times

Suffix	Type frequency (lemmas)	Token frequency
-ιον	150	28,645
-αριον	21	1,237
-ιδιον	16	1,094
-υδιον	4	78
-ισκη	1	135
-ισκος	1	28
-υδριον	1	28
-αδιον	1	20

After excluding these from the automatically extracted list, I linked each diminutive to the noun that it derives from. This was done in two ways: by data automatically extracted from the LSJ lexicon (Jones et al. 1996) (e.g. for παιδίον the dictionary entry of LSJ states 'Dim. of παῖς') and with a rule-based approach, where for all lemmas present in the knowledge database of Morpheus (Crane 1991) a diminutive form was generated on the basis of their inflectional class (e.g. since Morpheus lists the lemma νόμισμα with the inflectional class *ma_matos*, we know that the diminutive form would be νομισμάτιον) and afterwards the diminutives in the papyrus corpus were matched against this list. The latter approach would sometimes yield more than one option (e.g. πεδίον could theoretically be from πέδη or πέδον), and so the LSJ list was given priority when it contained a base word for the given diminutive word. Next, I manually checked the list using several scholarly sources (Beekes 2009; Lühr and Balles 2008; Petersen 1910) to make sure that the link between diminutive and base word was valid.

7.3.2 Isolating forms with a diminutive meaning

As discussed in Section 7.2, the 'diminutive' suffixes in Greek typically have a wide range of meanings. Importantly, for the diminutives that were called 'type A' the suffix can completely alter the denotational meaning of a word. For example, the most common diminutive word in the database I compiled was κεράτιον (accounting for more than one third of all ιον-diminutives), which typically is a measure of weight/price in papyri rather than meaning 'little horn' (as we would expect for a diminutive of κέρας). To retrieve the diminutives that are roughly interchangeable with their base form (see Section 7.2), it is necessary to isolate these instances from the list of diminutives.

There are two ways to do so: (a) by investigating all the secondary literature that is available on the target diminutives on this list or (b) by careful close reading of the corpus examples of all the target diminutives. While both options are very time-intensive, the former option is also rather problematic, given that most sources on diminutives are biased towards literary Greek and may not take all the papyrological evidence into account or display it very prominently (for example, the first meaning that the LSJ dictionary cites for κεράτιον is, in fact, 'small horn'). Ideally then, the latter option should be able to be automated,

so that one could take the specific contexts these words appear in in the papyrus corpus into account while saving time.

To make things more concrete, such a method should be able to distinguish diminutives that are more or less interchangeable from the ones where the suffix drastically changes the denotational meaning from the word. Luckily, such a method exists, viz. the broad family of techniques called 'distributional semantic models', which have already been applied to Greek numerous times (e.g. Keersmaekers and Speelman 2023; Rodda, Probert, and McGillivray 2019; Stopponi et al. 2023). These techniques represent each word as a vector describing the contexts in which the respective word occurs. Words that occur in similar contexts will have (mathematically) similar vectors, and are assumed to have similar meanings. While this chapter does not discuss the technicalities of calculating such word vectors,[5] it is important to note that this is to a large extent an automated way to do traditional philological analysis: just as we know that κεράτιον refers to money because it is often combined with ἀποδίδωμι, χρεωστέω, etc., such models determine κεράτιον's similarity with other money terms and its dissimilarity with κέρας based on the contexts it occurs in.

Concretely, I created word vectors for all the diminutives and their base words occurring in the papyrus corpus based on the contexts that they occur in. One should note that these models typically require very large amounts of data: the papyrus corpus of 4.5 million tokens is rather small in this respect, and problems with regard to data scarcity are discussed in what follows. Next, I measured the mathematical distance between diminutive and base word using the cosine similarity measure (see Turney and Pantel 2010, 160–61). In case the diminutive and base word occur in similar contexts, this similarity will be high, while it will be low when the contexts are very different.

An example of the results of this method, containing ten words chosen at random, is shown in Figure 7.1. Some diminutives that are judged highly interchangeable with their base form by the algorithm are δακτυλίδιον '(small) ring' and ἐπιστόλιον '(small) letter', which seems to correspond to their use in the papyri.[6] Some diminutives that occur in very different contexts from their base form are the aforementioned κεράτιον as well as κάδιον. Since κάδιον and κάδος both seem to refer to jars in papyrus texts, it seems more puzzling why the algorithm judged them to be very dissimilar. However, since κάδιον

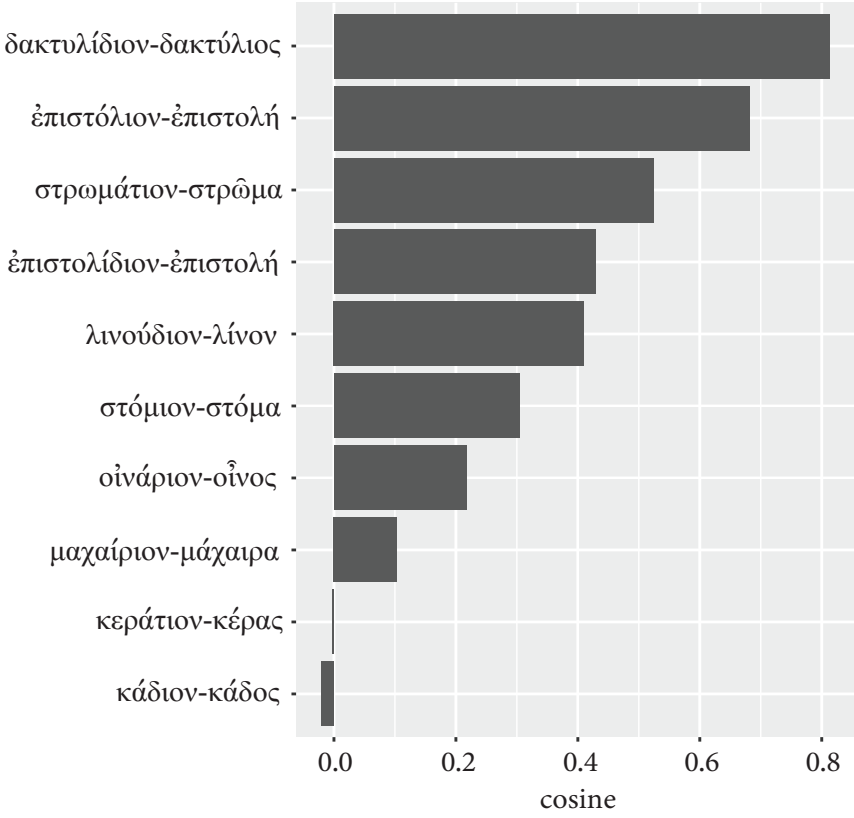


Figure 7.1: Cosine similarities between diminutives and their base forms, for a sample of ten words chosen at random

only occurs 15 times in the papyri, this is likely simply related to data scarcity.[7]

To further evaluate how well the algorithm performs, I looked at the ten words that were judged most similar to their base forms and the ten words that were judged the most dissimilar (while being fully aware that this is an extremely small sample). The most similar words were δακτυλίδιον 'ring' (vs δακτύλιος), ἀρνίον 'lamb' (vs ἀρήν), νομισμάτιον 'solidus' (vs νόμισμα), ἐρίφιον 'kid' (vs ἔριφος), χιτώνιον 'dress' (vs χιτών), ἐπιστόλιον 'letter' (vs ἐπιστολή), χρυσίον 'golden coin' (vs χρυσός), μαφόρτιον 'veil' (vs μαφόρτης), ληκύθιον 'flask' (vs λήκυθος), and τυρίον 'cheese' (vs τυρός). The only problematic case on

this list is χρυσίον, which means 'golden coin' instead of 'gold' as χρυσός does. However, they both often occur in transactional contexts, for example combined with words such ἀπόδοσις or ἀπαιτέω, which has likely caused their high vector similarity.

The most dissimilar words were δένδριον (vs δένδρον), κοράσιον (vs κόρη), πλινθίον (vs πλίνθος), κηπίον (vs κῆπος), ταξίδιον (vs τάξις), πλινθούργιον (vs πλινθουργός), ὀνάριον (vs ὄνος), κάδιον (vs κάδος), κεράτιον (vs κέρας), and στάθμιον (vs σταθμός). Judging by lexicons and the corpus evidence, the algorithm was right in detecting that πλινθίον and πλίνθος have very different referents (typically 'rectangle' vs 'brick'), as do ταξίδιον and τάξις ('journey' vs 'rank', or other meanings inherited from classical Greek), πλινθούργιον and πλινθουργός ('brickworks' vs 'brickmaker'), and κεράτιον and κέρας (see earlier in this section). κοράσιον and κόρη also have somewhat different referents (κοράσιον typically refers to 'slave girl', while κόρη simply means 'girl' or sometimes 'pupil').[8] On the other hand, κηπίον and κῆπος seem to be used roughly synonymously ('garden'), as do ὀνάριον and ὄνος ('donkey'), κάδιον and κάδος (see earlier in this section), and στάθμιον and σταθμός (a weight measure). Finally, for δένδριον it is difficult to say whether it truly has a different referent from δένδρον 'tree' or not: all 11 examples of δένδριον in the papyri occur in the same text (*Stud. Pal. 10 75*, TM 33215) in a similar formulaic context (ἐκκλησίας τοῦ *NAME* [e.g. Ἁγίου Ἀποστόλου] κοληφάνια *NUMBER* [e.g. ιβ] εἰς τέντρια *NUMBER*, where the editor has corrected τέντρια to δένδρια) where it is difficult to say what the meaning of δένδριον is (if τέντρια is actually from the lemma δένδριον to start with).

In general the formulaic and repetitive nature of the papyrus corpus complicates the use of distributional semantic models, since it can heavily skew the frequencies of the words with which our target word is frequently combined. This is especially problematic since the papyrus corpus is not a very big corpus, and low frequency has undoubtedly caused several of the words listed above to be judged dissimilar even though they are similar in meaning. Sometimes there is also a very big frequency discrepancy between diminutive and base word, such as ὀνάριον, which occurs only 12 times vs ὄνος, which occurs 2,820 times. These discrepancies are in themselves a big problem, given that the contexts of low-frequency words and high-frequency words tend to be very different.[9]

In statistical terms, if this small sample is representative, we can therefore say that this method has high precision (i.e. low amount of false positives) but lower recall (i.e. considerable amount of false negatives): while the top words with the highest cosine similarity do seem to be cases where the diminutive suffix does not alter the meaning of the word too much, in the bottom words there are also quite a few words where the meaning is not changed too much likewise. As the goal of this chapter is not, however, to give an exhaustive overview of all diminutive uses in the papyri but to retrieve a large dataset of diminutives that are more or less variable with their base words, these results were satisfying enough to proceed.

Concretely, I included every diminutive that had a cosine similarity of at least 0.5 (an arbitrarily chosen number) to its base form in this study. These were 20 words belonging to seven semantic fields, as listed in Table 7.2. I further manually filtered this list to the words in bold, by excluding some terms where there are still substantial meaning differences between base word and diminutive: αἰγίδιον is typically used for the young of a goat instead of a goat; χρυσίον was discussed above; ὀψαρίδιον seems to have a more specific meaning ('fish') than

Table 7.2: Diminutives retrieved as highly similar to their base form by the distributional models (diminutives in bold were included in this study)

Semantic field	Diminutives
Clothing	**δακτυλίδιον/δακτύλιος 'ring'**, **χιτώνιον/χιτών 'chiton'**, **μαφόρτιον/μαφόρτης 'veil'**, **πορφύριον/πορφύρα 'purple garment'**, **κοσμάριον/κόσμος 'jewel'**, **ἁλυσίδιον/ἅλυσις 'necklace'**
Animals	**ἀρνίον/ἀρήν 'lamb'**, **ἐρίφιον/ἔριφος 'kid'**, **ὀρνίθιον/ὄρνις** 'bird, chicken', αἰγίδιον/αἴξ 'kid' vs 'goat'
Money	**νομισμάτιον/νόμισμα 'solidus'**, χρυσίον/χρυσός 'golden coin' vs 'gold'
Household objects	**ληκύθιον/λήκυθος 'flask'**, **στρωμάτιον/στρῶμα 'mattress'**, **κλειδίον/κλείς** 'key'
Food	**τυρίον/τυρός 'cheese'**, ὀψαρίδιον/ὀψάριον (ὄψον) 'fish, delicacy'
People	παιδάριον/παῖς 'slave' vs 'child'
Writings	**ἐπιστόλιον/ἐπιστολή 'letter'**, βιβλίδιον/βιβλίον 'petition' vs 'document'

Table 7.3: Distribution of diminutives included in this study

Semantic field	% base forms	% diminutives	Standard deviation
Clothing	0.748 (541)	0.252 (182)	0.231
Animals	0.751 (377)	0.249 (125)	0.119
Money	0.804 (18389)	0.196 (4477)	NA
Household objects	0.841 (232)	0.159 (44)	0.105
Food	0.672 (133)	0.328 (65)	NA
Writings	0.891 (1685)	0.109 (207)	NA
Total	**0.807 (21357)**	**0.193 (5100)**	**0.184**

ὀψάριον ('delicacy') (additionally, ὀψάριον is also a diminutive, so it is rather problematic to include it in this list); παιδάριον mainly seems to be used with the meaning 'slave', while παῖς can also be used for children; and βιβλίδιον again has a more specific meaning ('petition') than βιβλίον (any kind of document).

After filtering out these words, 15 terms remained; some general statistics of these forms are listed in Table 7.3. Diminutives seem particularly frequent with clothing terms and animal terms (as well as food terms, but this category contains only one representative, namely τυρίον).[10] In general there is a lot of variation between words, as can be judged from the standard deviation of 18.4 per cent: the number of diminutives ranges from 11 per cent (στρωμάτιον and χιτώνιον) to 75 per cent (μαφόρτιον).

7.4. Analysing variation and change in the use of diminutive suffixes

To analyse which factors cause the use of the diminutive suffix in quasi-interchangeable situations, I added a number of possible explanatory factors to the dataset. Focusing on variables that were easy to add to the dataset in an automatic way, these were:

- semantic class (the six classes discussed in Section 7.3, viz. clothing, animals, money, (household) objects, food, writings);
- period (Ptolemaic, Roman, Byzantine);

- region where the text was written (Fayum, Lower Egypt, Upper Egypt, Western Desert, Eastern Desert, Western Coast, other Egypt, Palestine, (other) non-Egypt);
- text type (administrative, i.e. accounts, contracts, lists, and receipts; high-register, i.e. declarations, official letters, petitions, pronouncements, and reports; low-register, i.e. private and business letters; other);[11]
- material (papyrus, pottery, other material).

All the extralinguistic information was retrieved from the Trismegistos databases (Depauw and Gheldof 2014).

It is plausible that several of these factors interact: the importance of region may be different across time periods, or the semantic classes that have a high number of diminutives may be different for various text types, for example. A good tool for exploring the interaction between several variables are conditional inference trees (see e.g. Levshina 2015, 291–300). The aim of a conditional inference tree model is to divide a complex dataset into subparts according to a number of possible explanatory variables. This is done in the following way: we start from the full dataset A, containing the variable we are interested in (e.g. the use of diminutive vs base form) which can be divided according to a number of explanatory variables (e.g. period, text type, material). From these factors we select the factor with the lowest *p*-value (e.g. period), which means that the chance is highest that a dataset where period can explain differences in diminutive use differs statistically from a dataset where this is not the case. Subsequently, we split this dataset into two datasets B and C according to this factor (e.g. dataset B, containing all data from the Ptolemaic period, has the diminutive in 30 per cent of all cases, while dataset C, containing all data from the Roman and Byzantine period, has the diminutive in 70 per cent of all cases) and investigate if each of these sub-datasets can be further split according to the most significant factor for the respective dataset. This splitting process continues until no more statistically significant factors (according to a certain threshold, e.g. p=0.05, which means that there is a 5 per cent chance that the influence of the given factor is completely random) can be found.

It is important to be very careful applying conditional inference trees (or any statistical method in general) to this dataset, however. In particular, it is very noisy: unlike typical research in variational

linguistics, where a careful effort is undertaken to make sure that the variants that we are looking at are actually interchangeable, this dataset inevitably still contains several diminutives that are not interchangeable with their base form, as discussed in Section 7.2. This is the case because I did not check all the corpus examples: it is likely that the dataset still contains instances where the diminutive and base forms have different types of referents, because of, for example, polysemy (the base form might have more meanings in some cases than the diminutive) or because the diminutive suffix adds an additional nuance (e.g. smallness). Conditional inference trees are therefore primarily used to generate hypotheses rather than to test them (i.e. to find general prevalent patterns in the dataset that could be investigated with further research).

Concretely, using R (R Core Team 2022) package *partykit* (Hothorn and Zeileis 2015), I made the parameters of the conditional inference trees somewhat stricter than the default ones to filter out as much statistical noise as possible: I set the threshold for a split to p=0.01 (meaning that only one in 100 splits with a p-value of 0.01 is entirely due to chance, rather than one in 20), the minimum number of instances that should be present in a 'node' (meaning the sub-dataset that is created from the split) to 50 (instead of seven), and the maximum depth of the tree (meaning the times that a dataset is further split into a smaller dataset) to three, so as not to present too-complex patterns in the dataset that are difficult to interpret. For the sake of clarity, I made separate trees for each period, so as to easily see diachronic changes in the dataset, although a tree for all three periods will also be discussed.

Figure 7.2 presents the conditional inference tree for the Ptolemaic period. It is rather simple and can be read as follows: for the whole dataset (1,211 instances) the rate of diminutives is somewhat low (only in 5 per cent of all cases). There is a genre contrast, however, with diminutives being much more present in low-register texts (in 12 per cent of all cases) than in high-register and administrative cases (only 3 per cent). For both high-register/administrative texts and low-register texts, no further statistically significant factors can be found to split the dataset further.

Figure 7.3 presents the conditional inference tree for the Roman period. In general, there is a much higher rate of diminutives than in the Ptolemaic period (23 per cent of all cases). There are also semantic contrasts: diminutives are more present with words referring to

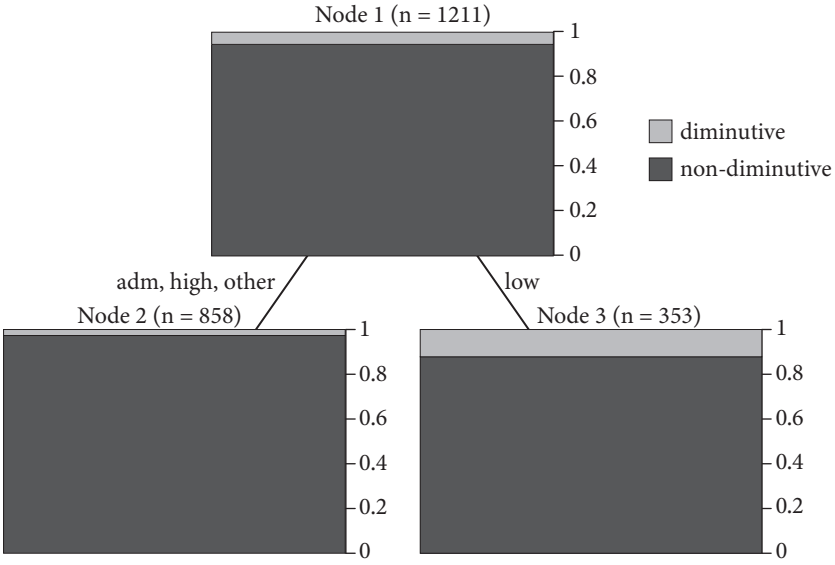


Figure 7.2: Conditional inference tree of diminutive use in the Ptolemaic period

animals, clothing, food (τυρίον), and money (νομισμάτιον) than to household objects and writings (ἐπιστόλιον). For both of these classes, there are genre contrasts, with diminutives again being more prevalent in low-register texts than high-register and administrative texts.

There are also regional variations, although they are a little different for the two semantic groups. For the diminutives referring to animals, clothing, food, and money, the most significant regional differences are between the Western Desert (66 per cent diminutives) and the other regions (28 per cent diminutives). Inspecting the data in more detail, most of the data points for the Western Desert are from the fourth century CE (65/71 or 92 per cent; the rest of the data points are from the third or the fourth century), as opposed to the other regions (513/1,438 or 36 per cent are from the fourth century); a large number of data points are from the village of Kellis (57/71 or 80 per cent); and there are no high-register texts, so there might be temporal, local, and/or genre factors involved.[12]

For words referring to household objects or writings (ἐπιστόλιον), there is a (smaller) regional contrast between the Fayum and texts outside Egypt that show a larger rate of diminutives (37 per cent) in

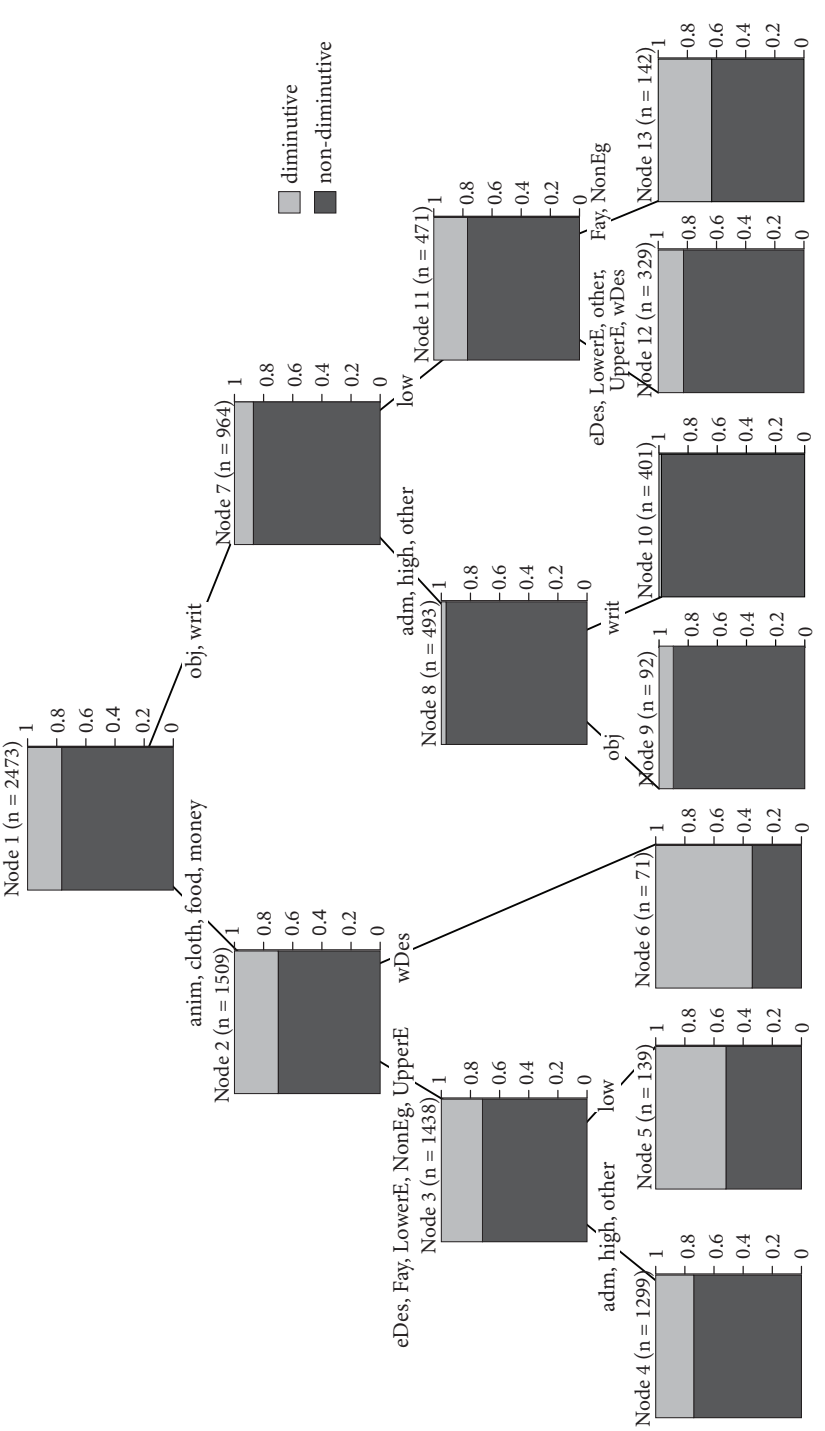


Figure 7.3: Conditional inference tree of diminutive use in the Roman period

low-register texts than the other regions (17 per cent). Finally, there is also a contrast between the diminutive ἐπιστόλιον, which is very rarely used instead of ἐπιστολή in administrative and high-register texts (only in 2 per cent of all cases), and the words referring to household objects, which show some diminutives there (in 10 per cent of all cases).

Figure 7.4 presents the conditional inference tree for the Byzantine period, showing a slightly lower rate of diminutives again (18 per cent), but overall comparable to the Roman period (where the rate of diminutives was 23 per cent). Firstly, one should note that this tree is extremely skewed toward the pair νόμισμα/νομισμάτιον: 20,612 (98 per cent) of the instances are this pair, as opposed to only 327 instances of the other pairs. Again, there are regional contrasts, primarily between the Fayum (as well as the Western Desert and Lower Egypt), where the high rate of diminutives (64 per cent) is even more striking than in the Roman period, and the other regions, with much lower rates of diminutives (13 per cent). Note that the number of data points for the Western Desert (3) and Lower Egypt (33) is rather low as compared to the Fayum (1808).

Within the Fayum, Lower Egypt, and the Western Desert, there are again genre contrasts, but they are rather surprising: this time the rate of diminutives is much higher in high-register and administrative texts (66 per cent) than in low-register texts (27 per cent). It is difficult to explain why this is the case, although it is important to note that this time the data of low-register/other texts is highly skewed towards one text, viz. P.Rain. Unterricht 80 (TM 39240), a school text containing 21 instances of the word νόμισμα, which are one third of all non-diminutive forms in this sub-dataset.

In the other regions, there are contrasts related to register as well: firstly, the rate of diminutives is much lower in texts written on papyrus (12 per cent) than on other materials (in particular pottery) (46 per cent), which again might point toward to diminutives being a low-register phenomenon. For texts written on papyrus, there is also a contrast between high-register texts (7 per cent diminutives) and administrative/low-register texts (13 per cent), although in general the rate of diminutives is rather low. Finally, for the texts written on pottery/other materials the main region causing a high rate of diminutives is Upper Egypt: the Western Coast does not show any diminutives (for the regions Eastern Desert, non-Egypt, and Palestine, there are no texts that are not written on papyrus). For both regions all instances are the

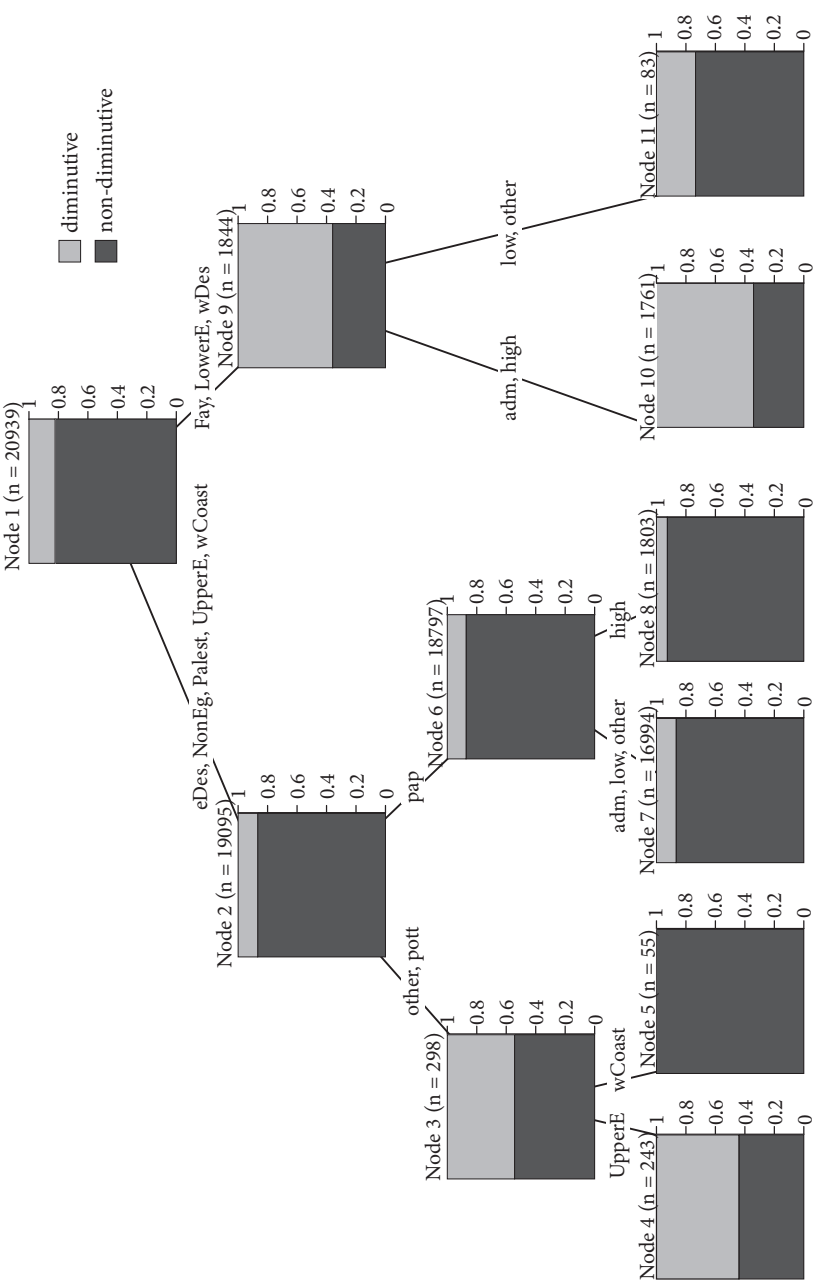


Figure 7.4: Conditional inference tree of diminutive use in the Byzantine period

pair νόμισμα/νομισμάτιον, but the text types are rather different: the data from the Western Coast are almost all payment orders (50/55 cases), all coming from the village of Abu Mena, while the data from Upper Egypt are mainly (accounting) lists and (tax) receipts (208/243 cases).

From the discussion in this section a diachronic picture emerges, but it is important to be careful. While the Ptolemaic period shows a lower rate of diminutives than the other periods, for example, it might be the case that certain text types or regions that also have a low rate of diminutives are over-represented in this period and there is no actual language change going on. Therefore I generated a final tree incorporating period as a factor, to get a good overview of the major change patterns. To compensate the addition of another factor, I increased the maximum depth of the tree to four (instead of three).

Figures 7.5a and 7.5b show the two branches of this final tree. It is immediately clear that the major diachronic change patterns can be situated in the Fayum region (as well as the Western Desert, Figure 7.5a), where there is an immediate split in terms of diachrony, while on the left branch of the tree (Figure 7.5b) no split occurs (at least not at the maximum level of depth specified).[13] In particular, in the former region, there is a strong increase in diminutives from the Ptolemaic period (5 per cent diminutives) to the Roman period (27 per cent diminutives) to the Byzantine period (64 per cent diminutives). For each of these periods, there are register contrasts following the patterns discussed above. Outside the Fayum, the rate of diminutives is rather low across time, and factors such as text type and material (relating to register), regional factors, and semantics seem to play a bigger role.

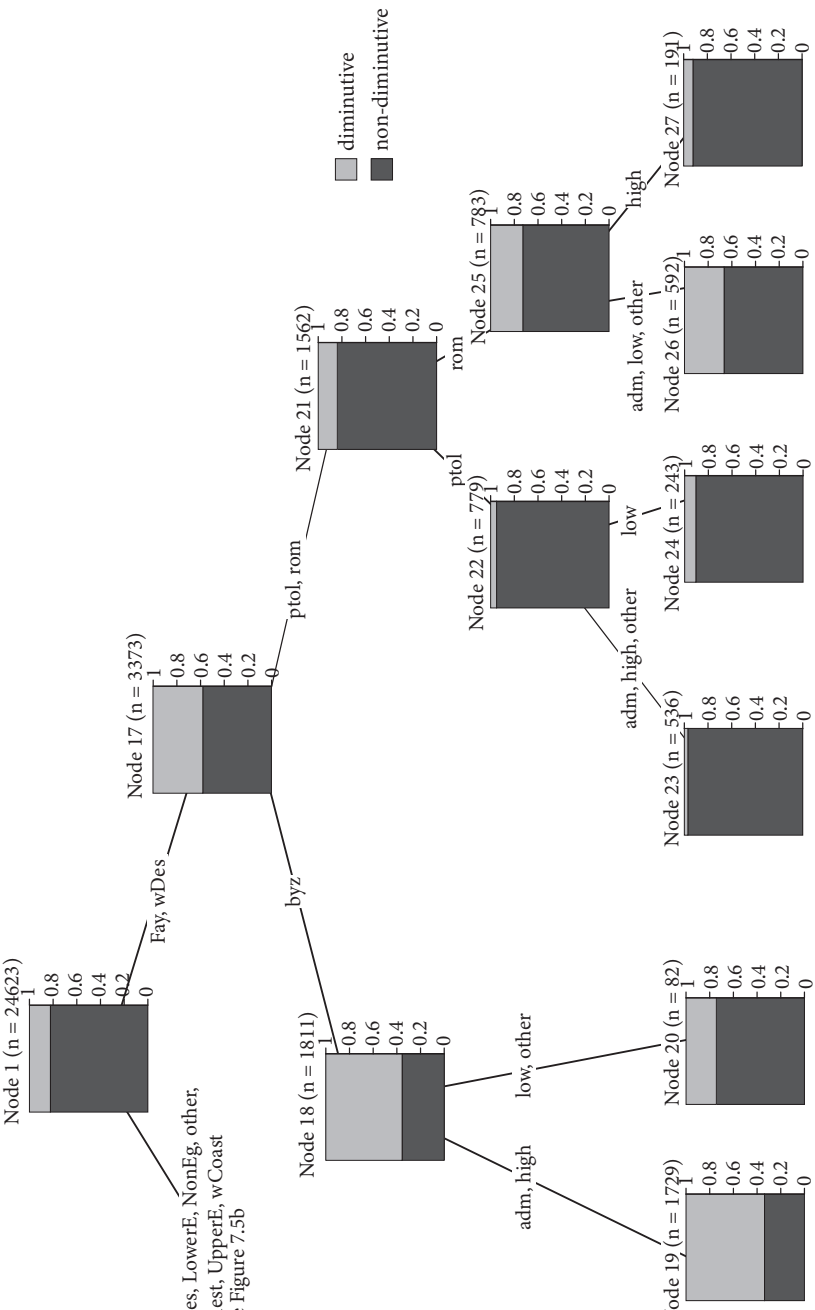


Figure 7.5a: Conditional inference tree of diminutive use in the Fayum region and the Western Desert in all three periods

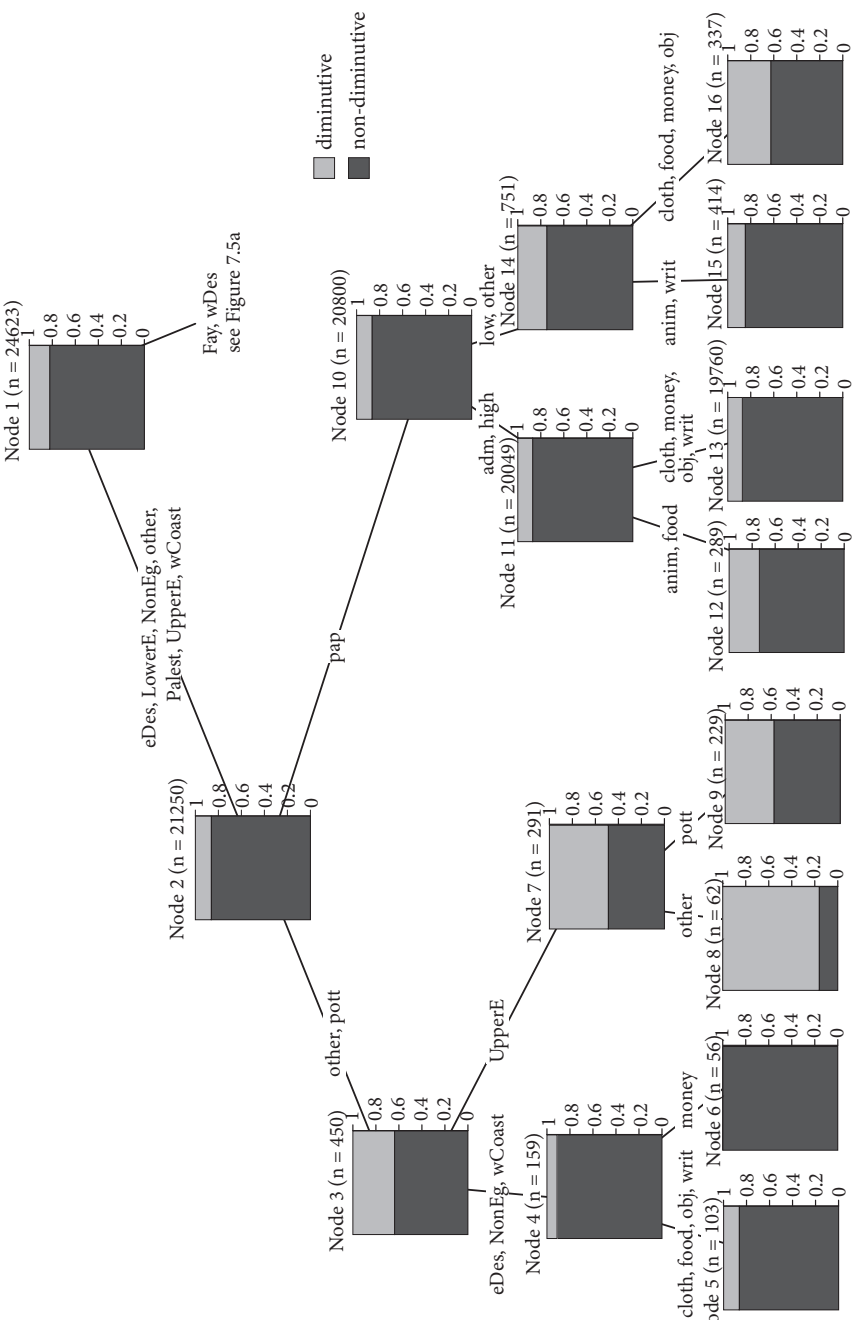


Figure 7.5b: Conditional inference tree of diminutive use in all regions except the Fayum and the Western Desert in all three periods

7.5 Discussion

This study is a first attempt to give a broad overview of the factors conditioning the use of the diminutive suffix in the Greek papyri, focusing on cases where the diminutive was broadly interchangeable with its base form. It uses quantitative techniques to analyse a large, automatically extracted dataset from the papyrus corpus. Although this dataset is rather messy, several major explanatory factors still emerged, including region, diachrony, and register.

As for region, diminutive suffixes seem to be particularly popular in the Fayum. It is also in this region that major change happens, showing an increasing number of diminutives from the Ptolemaic to the Roman to the Byzantine period, while in the rest of Egypt the situation seems to be more stable. Regarding register, diminutives seem to be in general a low-register phenomenon, except in Byzantine texts from the Fayum region (where diminutives are actually more frequent in high-register and administrative texts). This might point to the Fayum region being the 'leader' of the change, where diminutives are becoming more quickly interchangeable with non-diminutive forms, even in higher registers. Nevertheless, one should be careful not to overinterpret these results, especially because the data of the Byzantine period is highly centred around the pair νόμισμα/νομισμάτιον.

The semantics of the diminutive word also seems to play a role, although very definite conclusions are difficult to make. It is only in the decision tree of the Roman period that there was a high-level split between two categories of words (viz. words referring to animals, clothing, food, and money, which show a higher rate of diminutives than words referring to household objects and writings). Other than that, there are also some more specific patterns tied to specific regions and text types that are more difficult to interpret. Of course, it is not given that the semantic classes I used were the best classification (especially because some of them, in particular clothing, show very high standard deviations): it could be the case, for example, that there are very specific local preferences for the diminutive with certain lemmas rather than being tied to broader semantic groups. At any rate, the number of words included in each class was rather small (sometimes only one), and perhaps clearer semantic patterns would be revealed if a broader range of lemmas was selected.

Let us now reflect critically on the methodology used in this study, which consisted of two central techniques: the use of distributional semantic methods to distinguish 'interchangeable' diminutives from 'non-interchangeable' diminutives, and the use of quantitative methods (specifically conditional inference trees) to analyse the various factors conditioning diminutive use jointly.

As for distributional semantic methods, a first important methodological remark that must be made is the fact that for the papyrus corpus, distributional semantic methods are clearly at their limits, since it is a relatively small and highly peculiar corpus containing many repeated formulae that have a large impact on the context frequencies of our target words. A more targeted approach that takes the formulaic character of the papyrus corpus into account when calculating context frequencies would therefore be necessary (see also Levy, Goldberg, and Dagan 2015 on reducing the effect of frequency). Nevertheless, on a very rough level the method was able to isolate diminutives that are more or less interchangeable with their base form from those that are non-interchangeable (or where there is not enough corpus evidence that they are interchangeable). However, the word 'rough' is crucial here, since there were some problematic cases included in the results, for example, αἰγίδιον which is not interchangeable with αἴξ; the shared collocates of 'goat' and 'young of a goat' are likely too similar for the model to make an informed decision.

Although this problem was solved by checking the results manually, what is important to point out is that diminutives are often polysemous, while the method employed here simply looked at the similarity in meaning of the 'aggregate' of all base form uses, covering all their meanings, and likewise all diminutive uses. For example, κεράτιον was excluded because it was too dissimilar to κέρας, which seems justifiable, since κεράτιον is typically a measure of weight rather than referring to 'little horn', but it cannot be excluded that the 'little horn' meaning sometimes also occurs in papyri, as it sometimes does in literary texts (Jones et al. 1996). While for κεράτιον this seems like a stretch, for other words it is more plausible: for κοράσιον, for example, it might be the case that sometimes the intended meaning is 'little girl' rather than 'slave girl'. A possible solution might be to employ other quantitative models in the field of distributional semantics that try to model the meaning of a word in context rather than the aggregate meaning of a word (see Geeraerts et al. 2023, 65–87).

As for quantitative methods, caution needs to be employed when applying statistical methods to this dataset. There are two reasons for this. Firstly, the papyrus corpus is certainly not a balanced corpus, so that trends that seem to be caused by one factor (e.g. diachrony) might be caused by others (e.g. region, due to certain regions being over-represented in certain periods). Additionally, the formulaic language and large degree of repetition in general can skew certain patterns.

Secondly, this dataset in itself is rather messy, since it will inevitably contain many diminutives that are not strictly interchangeable with their base form. Of course, one could wonder to what extent a line can be drawn between 'interchangeable' diminutives and 'non-interchangeable' diminutives, or, in other words, in which cases we can establish that the diminutive changes the denotational meaning of a word. If a Greek speaker uses the diminutive suffix to convey a particular nuance such as smallness or positive/negative evaluation, the meaning is not very strongly altered (as opposed to cases such as κεράτιον)—and therefore also difficult to pick up by statistical methods—but it might not be strictly interchangeable with the base form. It is an open question how much this statistical 'noise' was drowned out by the large amount of data used in this study: if we assume that this has an effect on factors such as region (i.e. that regional patterns were distorted because the dataset contained too many 'non-variable' diminutives), it would also logically be the case that in certain regions people refer more often to smaller objects or objects that they evaluate positively/negatively, which seems a little implausible. Nevertheless, for some factors this seems more plausible; for example, it might be true that in lower text registers people more frequently evaluate objects positively/negatively than in higher registers. An alternative method might be to filter the dataset manually to include only cases that are truly interchangeable. However, this would be a rather subjective exercise, given that in most cases there are no overt contextual clues whether the diminutive adds a meaning nuance or not (even in cases where the meaning is more strongly altered, such as αἶξ 'goat' vs αἰγίδιον 'kid', it is often difficult to say whether αἰγίδιον actually refers to 'kid' or not).

At any rate, one could argue that the problems addressed here might call for an alternative methodology, such as a so-called 'close reading' approach. I would argue that the opposite is the case: given the complexity of the problem, it is implausible that the myriad of

factors explaining the use of diminutive suffixes in the papyri could be discerned with the human eye or that the large datasets needed to disentangle these factors properly can be analysed without statistical techniques. Although one could certainly have objections to the level of noise present in this dataset, it is precisely because of the automated way of creating the dataset that complex, large-scale patterns could be unearthed, while a fully manually annotated but smaller dataset might lead to data scarcity. Whether the statistical techniques employed here were able to filter out the noise in this dataset is a question that is open for further research.

One further limitation of this study is that I focused on factors that could be easily added to the data because they were already present in existing databases (viz. the Trismegistos project), while other sociolinguistic factors (e.g. gender, ethnic background, social class) might have a more important effect on the use of the diminutive suffix than the ones I examined. Nevertheless, it is clear that statistical techniques show promising potential for the complex topic of diminutives in the papyrus corpus.

Notes

1 These numbers were obtained by querying the duke-nlp corpus (Keersmaekers and Depauw 2022) and the GLAUx corpus (Keersmaekers 2021) for nouns of which the lemma ends in -ιον (or -ίον). This returns 75,173 results on 5,127,085 tokens in total for the papyri (relative frequency 14.7/1,000 words) vs 102,428 results on 24,200,537 tokens for the literary texts (relative frequency 4.2/1,000 words). Obviously this is only a very rough estimate of the frequency of diminutives, as will also be discussed in this chapter.

2 Sometimes these are called 'faded diminutives', but, confusingly, the term is employed for words of both type A and type C (e.g. Petersen 1910, 164–66).

3 Lühr and Balles (2008, viii) state that their work covers multiple time periods, although they do not specify whether this also includes papyri. If we compare their list with Palmer (1945), the only suffixes mentioned by Palmer that are not present in this list are -ασιον and -αριδιον, which would be covered by the general search for -ιον and -ιδιον and are extremely infrequent regardless.

4 -ακιον, -φιον, -ισκαριον, -ισκιον, and -υλλιον are also present, but none of the diminutives formed with these suffixes occurred more than ten times.

5 The models used are described in Keersmaekers and Speelman (2023), in particular the DepHeadChild model discussed there, with Singular Value Decomposition applied to reduce the vector to 100 dimensions.

6 E.g. P.Oxy. 10 1273 (TM 21791): ᾧ προσφέρει ἡ αὐτὴ ἐκδότις ἐπὶ τῇ αὐτῇ θυγατρὶ αὐτῆς καὶ γαμουμένῃ ἐν φερνῇ … **δακτυλίδιον** μεικρὸν τετά[ρ]ταις ἥμισυ 'to whom the same giver presents as dowry for the same daughter of her who is

marrying … a small **ring** of half a quarter'; SB 14 12200 (TM 30927): ἐκομισάμην τὸ **ἐπιστόλιον** ὃ ἔπεμψάς μοι τῇ κθ 'I received the **letter** that you sent to me on the 29th'.

7 Looking at the context words that κάδιον and κάδος co-occur with, there is not a large degree of overlap: the only shared ones are χάλκεος and καί, which are not very helpful for determining that they refer to similar objects.

8 However, the most likely reason that κοράσιον and κόρη were found to be very dissimilar is that forms of the lemmas κουρά/κουρεύς 'cutting'/'hairdresser' were often wrongly lemmatised as κόρη in the corpus.

9 Without going into too much technical detail, after calculating the contexts in which a given target word (i.e. in our case ὀνάριον and ὄνος) typically occurs, these context words are weighted by the information-theoretical measure PPMI, which makes sure that very frequent context words such as ὁ and καί do not dominate the results (see Keersmaekers and Speelman 2023 for more detail). However, this measure is heavily influenced by the frequency of the target words, so that the vectors of low-frequency words and high-frequency words will contain rather different numbers. There has been some effort to reduce this influence (Levy, Goldberg, and Dagan 2015), which could also have an effect on the papyrus corpus. At any rate, if a specific word has a low frequency, the chances are obviously much higher that most contexts it occurs in will not be descriptive of its meaning (especially if these contexts are very formulaic) than for high-frequency words, and so the problem remains.

10 However, the category of clothing terms has a standard deviation of 23 per cent: indeed, some of these words show very high rates of diminutives (e.g. μαφόρτιον, 42/56 times or 75 per cent; δακτυλίδιον, 30/60 times or 50 per cent), while others have very low rates (e.g. χιτώνιον, 43/384 times or 11 per cent; πορφύριον, 19/94 times or 20 per cent).

11 This is meant only as a proxy for the expected formality of a text, and one may argue about some of these classifications, for example whether texts such as petitions are truly all 'high-register' (see also Bentein 2015). Nevertheless, the plots below show that there are often stark differences between the texts that I labelled as 'high' and 'low' register. For a more detailed study, however, one may look at how diminutives behave in each of the text types individually instead of using an a priori classification of texts as 'high' and 'low' register as I did.

12 A plausible hypothesis might be that the distribution is skewed towards one particular text, but this is not the case: the 71 data points for the Western Desert contain 35 unique texts, and the text with the most data points only has six.

13 It is only after setting the maximum level of depth to six that a diachronic pattern emerges in the left branch of the tree: in administrative texts from these regions written on papyrus for words in the 'animal' and 'food' categories, there is a higher rate of diminutives in the Roman period (54 per cent, N=74), than in the Byzantine/Ptolemaic periods (23 per cent, N=111). However, this pattern is so peculiar that it can hardly be called a 'major' change pattern and might simply be a statistical quirk.

References

Amundsen, Leiv. 1965. 'Some Remarks on Greek Diminutives'. *Symbolae Osloenses* 40 (1): 5–16. https://doi.org/10.1080/00397676508590563.

Beekes, Robert. 2009. *Etymological Dictionary of Greek*. Leiden and Boston, MA: Brill.

Bentein, Klaas. 2015. 'The Greek Documentary Papyri as a Linguistically Heterogeneous Corpus: The Case of the "Katochoi of the Sarapeion" Archive'. *The Classical World* 108 (4): 461–84. https://doi.org/10.1353/clw.2015.0052.

Chantraine, Pierre. 1933. *La formation des noms en grec ancien*. Paris: Champion.

Crane, Gregory. 1991. 'Generating and Parsing Classical Greek'. *Literary and Linguistic Computing* 6 (4): 243–45. https://doi.org/10.1093/llc/6.4.243.

Depauw, Mark, and Tom Gheldof. 2014. 'Trismegistos. An Interdisciplinary Platform for Ancient World Texts and Related Information'. In *Theory and Practice of Digital Libraries: TPDL 2013 Selected Workshops*, edited by Łukasz Bolikowski, Vittore Casarosa, Paula Goodale, Nikos Houssos, Paolo Manghi, and Jochen Schirrwagen, 40–52. Communications in Computer and Information Science 416. Cham: Springer. https://doi.org/10.1007/978-3-319-08425-1_5.

Evans, Trevor V., and Dirk D. Obbink, eds. 2010. *The Language of the Papyri*. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780199237081.001.0001.

Geeraerts, Dirk, Dirk Speelman, Kris Heylen, Mariana Montes, Stefano De Pascale, Karlien Franco, and Michael Lang. 2023. *Lexical Variation and Change: A Distributional Semantic Approach*. Oxford and New York, NY: Oxford University Press. https://doi.org/10.1093/oso/9780198890676.001.0001.

Grandi, Nicola. 2014. 'Diminutives/Augmentatives (Syntax and Morphology)'. In *Encyclopedia of Ancient Greek Language and Linguistics*, edited by Vit Bubenik, Emilio Crespo, Chris Golston, Alexandra Lianeri, Silvia Luraghi, and Stephanos Matthaios, 488–98. Leiden: Brill. https://doi.org/10.1163/2214-448X_eagll_COM_00000108.

Gries, Stefan Th. 2017. 'Syntactic Alternation Research: Taking Stock and Some Suggestions for the Future'. *Belgian Journal of Linguistics* 31 (1): 8–29. https://doi.org/10.1075/bjl.00001.gri.

Hothorn, Torsten, and Achim Zeileis. 2015. 'partykit: A Modular Toolkit for Recursive Partytioning in R'. *Journal of Machine Learning Research* 16: 3905–9.

Jones, Henry Stuart, Henry George Liddell, Roderick MacKenzie, Robert Scott, and A. A. Thompson. 1996. *A Greek–English Lexicon*. New edition with new supplement. Oxford: Clarendon Press.

Keersmaekers, Alek. 2021. 'The GLAUx Corpus: Methodological Issues in Designing a Long-Term, Diverse, Multi-Layered Corpus of Ancient Greek'. In *Proceedings of the 2nd International Workshop on Computational Approaches to Historical Language Change 2021*, edited by Nina Tahmasebi, Adam Jatowt, Yang Xu, Simon Hengchen, Syrielle Montariol, and Haim Dubossarsky, 39–50. Online: Association for Computational Linguistics. https://doi.org/10.18653/v1/2021.lchange-1.6.

Keersmaekers, Alek, and Mark Depauw. 2022. 'Bringing Together Linguistics and Social History in Automated Text Analysis of Greek Papyri'. In *Digital Text Analysis of Greek and Latin Sources: Methods, Tools, Perspectives*, edited by Stelios Chronopoulos, Felix K. Maier, and Anna Novokhatko. *Classics@* 20 Special

Issue. Harvard: Center for Hellenic Studies. https://classics-at.chs.harvard.edu/bringing-together-linguistics-and-social-history-in-automated-text-analysis-of-greek-papyri/.

Keersmaekers, Alek, and Dirk Speelman. 2023. 'Applying Distributional Semantic Models to a Historical Corpus of a Highly Inflected Language: The Case of Ancient Greek'. *Glottometrics* 55: 17–43. https://doi.org/10.53482/2023_55_410.

Levshina, Natalia. 2015. *How to Do Linguistics with R: Data Exploration and Statistical Analysis*. Amsterdam and Philadelphia, PA: John Benjamins.

Levy, Omer, Yoav Goldberg, and Ido Dagan. 2015. 'Improving Distributional Similarity with Lessons Learned from Word Embeddings'. *Transactions of the Association for Computational Linguistics* 3: 211–25. https://doi.org/10.1162/tacl_a_00134.

Lühr, Rosemarie, and Irene Balles. 2008. *Nominale Wortbildung des Indogermanischen in Grundzügen. Band 1: Latein, Altgriechisch*. Hamburg: Kovač.

Meluzzi, Chiara. 2017. 'Diminutives in Ancient Greek: Intensification and Subjectivity'. In *Exploring Intensification*, edited by Maria Napoli and Miriam Ravetto. Amsterdam and Philadelphia, PA: John Benjamins. https://doi.org/10.1075/slcs.189.07mel.

Palmer, Leonard R. 1945. *A Grammar of the Post-Ptolemaic Papyri*. Vol. 1: Accidence and Word-Formation. Part 1: The Suffixes. London: Oxford University Press.

Petersen, Walter. 1910. *Greek Diminutives in -ion: A Study in Semantics*. Weimar: R. Wagner Sohn.

Petersen, Walter. 1913. *The Greek Diminutive Suffix -isko, -iskè*. New Haven, CT: Yale University Press.

Pijpops, Dirk. 2020. 'What Is an Alternation? Six Answers'. *Belgian Journal of Linguistics* 34 (December): 283–94. https://doi.org/10.1075/bjl.00053.pij.

R Core Team. 2022. 'R: A Language and Environment for Statistical Computing'. Vienna: R Foundation for Statistical Computing. www.R-project.org/.

Rodda, Martina A., Philomen Probert, and Barbara McGillivray. 2019. 'Vector Space Models of Ancient Greek Word Meaning, and a Case Study on Homer'. *Traitement Automatique des Langues* 60 (3): 63–87.

Sommerstein, Alan H. 1995. 'The Language of Athenian Women'. In *Lo spettacolo delle voci*, edited by Francesco De Martino and Alan H. Sommerstein, 61–85. Bari: Levante.

Stopponi, Silvia, Nilo Pedrazzini, Saskia Peels, Barbara McGillivray, and Malvina Nissim. 2023. 'Evaluation of Distributional Semantic Models of Ancient Greek: Preliminary Results and a Road Map for Future Work'. In *Proceedings of the Ancient Language Processing Workshop*, edited by Adam Anderson, Shai Gordin, Bin Li, Yudong Liu, and Marco C. Passarotti, 49–58. Varna: INCOMA.

Swanson, Donald C. 1958. 'Diminutives in the Greek New Testament'. *Journal of Biblical Literature* 77 (2): 134–51. https://doi.org/10.2307/3264595.

Turney, Peter D., and Patrick Pantel. 2010. 'From Frequency to Meaning: Vector Space Models of Semantics'. *Journal of Artificial Intelligence Research* 37: 141–88.

Watt, Jonathan M. 2013. 'Diminutive Suffixes in the Greek New Testament: A Cross-Linguistic Study'. *Biblical and Ancient Greek Linguistics* 2: 29–74.