Dating Greek Papyri with Text Regression

John Pavlopoulos[♠♣], Maria Konstantinidou[◇], Isabelle Marthot-Santaniello[°], Holger Essler[‡], Asimina Paparigopoulou[♠]

[♠]Department of Informatics, Athens University of Economics and Business, Greece
{annis,asimina}@aueb.gr
[♣]Department of Computer and Systems Sciences, Stockholm University, Sweden
[◇] Democritus University of Thrace
mkonst@helit.duth.gr
[°] University of Basel, Switzerland
i.marthot-santaniello@unibas.ch
[‡] Ca'Foscari University of Venice, Italy
holger.essler@unive.it

Abstract

Dating Greek papyri accurately is crucial not only to edit their texts but also to understand numerous other aspects of ancient writing, document and book production and circulation, as well as various other aspects of administration, everyday life and intellectual history of antiquity. Although a substantial number of Greek papyri documents bear a date or other conclusive data as to their chronological placement, an even larger number can only be dated tentatively or in approximation, due to the lack of decisive evidence. By creating a dataset of 389 transcriptions of documentary Greek papyri, we train 389 regression models and we predict a date for the papyri with an average MAE of 54 years and an MSE of 1.17, outperforming image classifiers and other baselines. Last, we release date estimations for 159 manuscripts, for which only the upper limit is known.

1 Introduction

Ancient textual artefacts are arguably the richest source of information on the ancient world. In the Graeco-Roman world and particularly in its Greek-speaking part, the most extensive coeval texts come from inscriptions and papyri. The latter is a collective term used for all ancient manuscripts, regardless of their writing material which, apart from papyrus, may be parchment, pottery, wood, and others. To correctly evaluate and make good use of these texts, we need to determine their date, provenance and historical context of their production and use. As far as dating is concerned, the value of the relevant evidence provided by the artefacts themselves varies considerably, ranging from a direct date in the text (following, of course, the calendar and dating system of the respective historical period) to no evidence at all. In between, there are

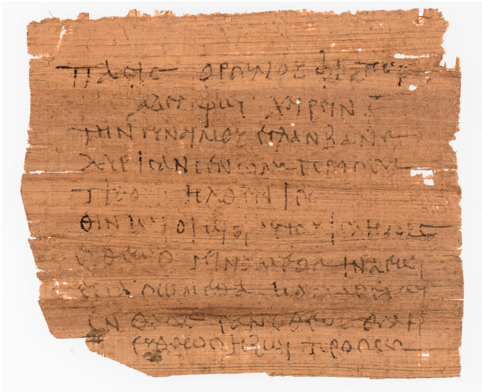

Figure 1: Papyrus 'P. Basel 2 15'. Credit: University of Basel. https://papyri.info/ddbdp/p.bas;2;15

texts containing references to known historical figures and events of a certain period, papyri which have been found next to other objects that can be dated, or other indirect evidence. The presence or absence of a date depends on the type of text preserved on the papyrus and its use through time, as well as on its state of conservation. Just like in modern times, it is much more likely to include a date in an official letter than in a page torn from a novel book. At the same time, it is more probable to find a date in a fully surviving letter than in a damaged one missing, for instance, the upper part of the first page.

Greek papyri, which mostly survive in fragments, are divided into two broad categories: books (literary and sub-literary papyri) and documents of all kinds (documentary papyri). The former ones never carry a date, whereas the latter often do, albeit not always unambiguously convertible by modern scholars. Most importantly for our study, literary papyri contain copies of works authored many years (often centuries) before the production of the

Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics
Volume 1: Long Papers, pages 10001–10013

actual manuscripts. On the other hand, documentary texts were usually written down as they were composed or shortly after that, making the content of their texts contemporary to their writing style or script. Therefore, any temporal indication in the text is also dating evidence regarding the production of the document. Even when there is no direct date in the text (e.g. Figure 1), documentary papyri can be dated securely sometimes within a short time-frame, because they may refer to known historical events or concern people known through other sources to have lived at a particular time.

When neither direct or indirect dating is possible, papyrologists resort to palaeography, the study of the script. In palaeography, particular writing styles are associated with certain chronological periods. Therefore, similar writing styles point to similar dates (Mazza, 2019). Securely dated specimens are used as a guide to chronologically place the undated ones. Growing criticism on the subjectivity of palaeographical dating (Mazza, 2019; Choat, 2019; Nongbri, 2019, 2014) highlights the need for more reliable methods. Recent efforts for computational dating of historical manuscripts are based on the script rather than the text and, although they consider various languages, they disregard Greek (Omayio et al., 2022).

In this study we focus on computational dating of Greek documentary papyri based on their transcriptions, contributing in the following three ways:

1. We present and publicly release a machine-actionable dataset of 389 documentary Greek papyri, containing texts of various aspects of daily life (e.g. contracts, receipts, letters).

2. We draw the baseline in text regression for the tasks of dating experimenting with Monte Carlo and leave one out cross validation.

3. We apply a committee of regressors to three papyri, which present different types of dating challenges, and on 159 manuscripts for which only the upper date limit is known.

This approach does not apply to literary papyri and our research involves solely documents. Apart from their texts being contemporary with the actual manuscripts (by dating the text, we date the papyrus), nonliterary papyri also include vastly more numerous objectively dated specimens than literary ones. Specific dates on our training set also allow for more accurate (narrower date-spans) predictions by our models.

2 Related Work

Dating historical documents with computational means has been studied for many languages (Baledent et al., 2020; Dhali et al., 2020; Li et al., 2015; Hamid et al., 2019; Adam et al., 2018). However, very limited work has been done for Greek and no published work at all has focused on Greek papyri. The only work to our knowledge is Ithaca, a Transformer trained on ancient Greek inscriptions performing text restoration, geographical attribution, and dating (Assael et al., 2022). Ithaca has achieved an error of 0.29 centuries in dating epigraphs. This result is by far better than an onomastic baseline using the known distribution of Greek personal names to infer the date, which scored 1.44. Inscriptions differ from papyri in many aspects (such as the genre, the length, and their geographical distribution), but in principle, this system is applicable to our data and was therefore used as a baseline. Below, given the absence of dating studies for Greek, we summarise work for other languages.

The studied languages are Latin (Baledent et al., 2020; Wahlberg et al., 2016, 2015), Hebrew (Dhali et al., 2020), Dutch (Hamid et al., 2019, 2018; He et al., 2014, 2016b,a), Arabic (Adam et al., 2018), Swedish (Wahlberg et al., 2016, 2015), French (Baledent et al., 2020) and English (Li et al., 2015; Rastas et al., 2022). A collection of 595 Dead Sea Scrolls, in Aramaic script, was the dataset with the oldest manuscripts, dated from 250 to 135 BCE, and the only one so far concerning texts written on papyri (Dhali et al., 2020). The rest of the datasets comprised more data, ranging from less than five (Adam et al., 2018) to more than ten thousand manuscripts (Wahlberg et al., 2015) or more (Rastas et al., 2022), while the one with the most recent manuscripts comprises historical English-language documents (Li et al., 2015), printed between the 15th and 19th CE.

The employed methods usually were standard machine learning methods, such as KNN (Adam et al., 2018), decision trees (Baledent et al., 2020), random forests (Baledent et al., 2020) and support vector machines (Hamid et al., 2019; Dhali et al., 2020; He et al., 2014, 2016b,a). Textural features, such as Gabor filters, Uniform Local Binary Patterns and Histogram of Local Binary Patterns are extracted and then fed to the classifiers (Hamid et al., 2018). The writing style evolution, however, has also been used as an intermediate step (Dhali et al., 2020; Adam et al., 2018). In this case, the periods are first aligned with specific writing styles. Then, any new manuscript is dated based on the detected style.

Pre-trained convolutional neural networks have been used to extract features, which are passed to a classifier or regressor (Hamid et al., 2019; Wahlberg et al., 2016), or used in combination with text features extracted with optical character recognition methods (Li et al., 2015). Transfer learning has been reported to lead to human performance (Wahlberg et al., 2016). This was deemed to be the most promising direction for the present study on Greek manuscripts, and was, hence, employed.

3 Data

Our dataset, which we release publicly,[1] comprises the transcriptions of 389 manuscripts, dated from the 3rd century BCE to the 7th century CE, originating from Greco-Roman Egypt (with a few exceptions from the Near-East).

3.1 The source

The dataset was compiled mainly from PAPYRI.INFO.[2] The documents in its collections set a reliable point of reference for scholars who aspire to study the evolution of ancient manuscripts in time. These collections incorporate full transcriptions and references to scholarly editions of the papyri, as well as a set of metadata that can also assist in dating (e.g. provenance).

3.2 The scripts and the language

Nonliterary papyri in Greek from the 3rd c. BCE to the 7th c. CE are written in a great variety of cursive hands (Harrauer, 2010), posing an extra challenge for image classification methods and calling for other approaches. The language of the papyri, Greek of the Ptolemaic, Roman and early Byzantine periods, reflects the diversity and the diachronic changes of the Greek-speaking communities in Egypt, which is the provenance of most of our specimens.

3.3 The ground truth

The date of a manuscript may be found in different forms. It can be an exact date, a range of years, a starting date (not before that date), or an ending date (not after that date), or two-three alternative dates. Our dataset has been curated so that dating applies at the level of the quarter of the century, by considering manuscripts dated exactly or with a period ranging within that quarter. We did not consider manuscripts that were dated only before or after a specific date.

3.4 Data collection

Our first dataset comprised 400 manuscripts, 40 samples per century. Our initial pool consisted of 77,040 items and we opted for ones that satisfy the following conditions:

- The transcriptions must be available in machine actionable form.
- The papyri must contain documents (not works of literature) to ensure that text and papyrus are contemporary.[3]
- The papyri must be securely and accurately dated. Many papyri do not carry a date and are, therefore, dated with subjective criteria or with a large date span (e.g. 1st-2ndCE).
- The image is available, to allow image-based dating and potentially jointly from different modalities: text and image.

Given these limitations, it was the 7thCE that dictated the size per century of a balanced dataset, since there are not more than 40 securely dated papyri from 7thCE. For each of these records, the text was retrieved afterwards from PAPYRI.INFO by parsing the respective XML files. We discarded records whose extracted text was less than ten characters, which resulted in our final 389 records. From these records, we extracted the entire text from one side of the papyrus (the side that had more text than the other). In the few cases of papyri with more than one fragment, we only included the first one. This decision was based on weighing the benefit of avoiding a considerable amount of noise during automatic parsing against eliminating a portion of text, in a dataset whose nature is by definition fragmentary.

3.5 Normalisation

The transcribed text comprises a variety of characters and symbols. We preprocessed the data by lowercasing and normalising the text (see Table 1). We

[1] https://github.com/ipavlopoulos/padoc

[2] https://papyri.info/

[3] Literary papyri are written on a certain date but may transmit a work of literature composed centuries earlier and there is no point in attempting to date the text (the date of composition is already known in most cases).

(ο)	‘ὅ’, ‘ὃ’, ‘ό’, ‘ό’, ‘ὄ’, ‘ὁ’, ‘ὀ’, ‘ὸ’
(α)	‘ἂ’, ‘ᾴ’, ‘ἃ’, ‘ᾅ’, ‘ά’, ‘ἁ’, ‘ἄ’, ‘ᾶ’, ‘ἀ’, ‘ᾷ’, ‘ὰ’, ‘ά’, ‘ἅ’
(η)	‘ἥ’, ‘ἣ’, ‘ῃ’, ‘ἢ’, ‘ἠ’, ‘ή’, ‘ἦ’, ‘ἤ’, ‘ῆ’, ‘ἧ’, ‘ᾗ’, ‘ᾕ’, ‘ῇ’, ‘ἡ’, ‘ῄ’, ‘ᾑ’, ‘ὴ’, ‘ή’
(ι)	‘ί’, ‘ῖ’, ‘ἷ’, ‘ἱ’, ‘ἶ’, ‘ῒ’, ‘ί’, ‘ἰ’, ‘ὶ’, ‘ΐ’, ‘ἴ’, ‘ι’, ‘ϊ’, ‘ἳ’, ‘ἵ’
(ε)	‘έ’, ‘ὲ’, ‘έ’, ‘ἕ’, ‘ἐ’, ‘ἔ’, ‘ἑ’, ‘ε’
(υ)	‘ὗ’, ‘ύ’, ‘ῦ’, ‘ύ’, ‘ὕ’, ‘ὓ’, ‘ὺ’, ‘ὑ’, ‘ὔ’, ‘ὐ’, ‘ὖ’
(ρ)	‘ῥ’, ‘ῤ’
(ω)	‘ώ’, ‘ῶ’, ‘ὧ’, ‘ᾤ’, ‘ὠ’, ‘ὦ’, ‘ῳ’, ‘ῷ’, ‘ὣ’, ‘ὥ’, ‘ώ’, ‘ᾧ’, ‘ὤ’, ‘ᾦ’, ‘ὤ’, ‘ᾠ’, ‘ὼ’, ‘ὡ’, ‘ῴ’
(σ)	‘σ’, ‘ς’

Table 1: Normalisation rules of characters in the dataset, all characters on the right have been replaced by the character on the left.

also discarded any character besides the 24 Greek letters, also removing white space and all punctuation marks. We did not eliminate the editors' corrections and supplements nor edit otherwise the data, which often led to duplicate words with alternative orthography (original and normalisation).

The transcriptions available are not diplomatic (reflecting exactly what is written) but normalised according to modern conventions, for example as far as punctuation and word separation (or sometimes spelling) are concerned. Therefore, we chose to disregard these conventions, because they do not represent data present in our sources, but normalisation on the papyrologists' part for the purpose of scholarly editions.

To provide some more concrete examples, there is no capitalization of proper names or initial words in sentences in papyri. Punctuation is very scarce and sometimes completely absent. Diacritics are not meaningless, but they are extremely rare in documentary papyri (i.e., except diaresis which is used in a different way than modern conventions, to mark iota and upsilon as the first letter of a word). Breathings and accents are marked inconsistently (if at all) by different scribes. Hence, removing diacritics leads to inclusion and can help avoid multiple variations of what is in fact the same word. Regarding spelling, we kept both the original and the corrected form (if provided by the editors), because spelling mistakes reflect language evolution.

3.6 Exploratory analysis

The overall text length per quarter of century varies over time, as can be seen in Figure 2. Although we have selected an equal number of manuscripts per century (§3.4), the number of lines within each manuscript varies, and so does the line length. Furthermore, within a century, manuscripts of a specific quarter of a century may be more frequent due to random discoveries, as is the case of 7thCE, where the first quarter holds most of the support, a discrepancy deriving from the reduced number of dated papyri in this century overall.

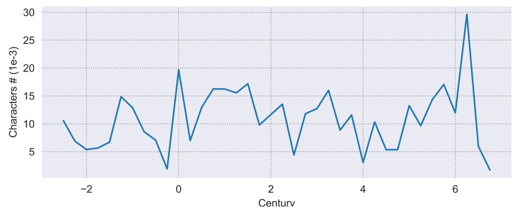

Figure 2: Text length (total number of characters divided by 1e3) in the manuscript transcriptions over time.

The most frequent character in our dataset is ‘α’ (35,101 occurrences), followed by ‘ο’ (33,176), ‘ι’ (30,151), and ‘ε’ (25,116). On the other hand, the least common are ‘β’ (2520), ‘ξ’ (1210), ‘ζ’ (379), and ‘ψ’ (334). These figures are coherent with general frequencies of letters in Ancient and Modern Greek (Mikros et al., 2005).

In order to assess the quality of the ground truth, we employed the Callimachus' Conservation number (CCN),[4] which provides an educated estimation of the preservation and legibility of a papyrus. The lowest score is 0 and the highest score (i.e., 1) indicates readability and 'perfect' conservation of the text. The status of the conservation of a papyrus affects the quality of the transcription, indicating the amount of text that has not been recorded in the transcriptions (or recorded with some level of uncertainty) because of the material state of preservation of the manuscripts. Damage in papyri could affect as little as one or two letters (or even none), to as much as several lines and whole parts of the

[4] https://glg.csic.es/Callimachus/Concordancia_Callimachus.html

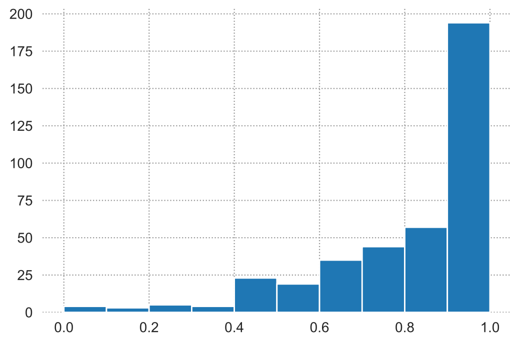

Figure 3: Callimachus' number histogram on our data.

papyrus sheet. As is shown in Figure 3, our dataset comprises mostly high-quality preservation scores.

4 Methodology

To estimate the date of production of manuscripts, we opted for text regression, taking advantage of the continuous target objective. Statistical validity was established with 5-fold Monte Carlo cross-validation. The best regression method was used to form a committee of models, which were applied on unseen data in order to analyse the predictions.

4.1 Benchmarking

We performed Monte Carlo cross-validation, by sampling 90% for training, 10% for validation, and then re-sampling with replacement five times. We report the mean absolute error (MAE), the mean squared error (MSE), and the explained variance (R^2). Besides the average results across folds, we also report the best score achieved per metric.

4.2 Regression methods

Fernández-Delgado et al. (2019) surveyed 77 regression methods and undertook an experimental analysis on 83 datasets. Regression with extremely randomised trees achieved the best R^2 in many datasets but gradient boosting and random forests were also found to have a promising performance. Following these findings, we opted for extremely randomised trees, random forests, gradient boosting, and linear regression for our experiments.[5]

Extremely randomised trees (XTrees) is a tree-based ensemble, created with the Extra-Trees algorithm (Geurts et al., 2006). Although simple in nature, it is both accurate and efficient Fernández-Delgado et al. (2019). Compared to other ensembles that use decision trees, XTrees splits the nodes of the tree by choosing randomly cut-off points and the trees grow by using the whole sample to learn instead of bootstrapping.

4.3 The Committee

Using the best-performing regression method out of the ones examined, we performed leave one out cross-validation, which allowed an evaluation using the whole dataset. Furthermore, it yielded as many regressors as the data points, which in our case is 389. We used these models to form a committee and date unseen papyri (further discussed in §6).

5 Empirical analysis

This section presents our experimental results using regression on textual features to date Greek manuscripts. First, we present preliminary experiments and then we analyse the experimental findings from our regression analysis.

5.1 Preliminary experiments

Preliminary experiments comprised image classification (Hamid et al., 2018), text classification with Transformers trained on another domain (Assael et al., 2022), and transferring learning from large language models (Koutsikakis et al., 2020).

Image classification was used prior to using transcribed text as our input, experimenting with using the documents' images (Hamid et al., 2018; Wahlberg et al., 2016; Paparigopoulou et al., 2022). Vanilla convolutional neural networks were outperformed by a pre-trained one (Tan and Le, 2019), fine-tuned for our dating task. Our estimated MAE, however, was consistently more than a hundred years (Table 2), hence we opted for textual input.

Ithaca was presented by Assael et al. (2022), consisting of a Transformer that is trained not only in dating but also in text restoration and geographical attribution. Ithaca has achieved an error of 0.29 centuries in dating inscriptions, which is by far better than an onomastics baseline (error of 144 years). By using the open-access web interface,[6] we scored all our preprocessed texts,[7] registering a MAE of approx. one century by using the maximum decade predicted or the average of the distribution (Table 2). The difference from the published result

[5] For all evaluation measures and algorithms we used the implementations of SCIKIT-LEARN.

[6] `https://ithaca.deepmind.com`

[7] We kept white space, to follow their standard.

possibly stems from the fact that this is a model trained and focused on inscriptions, not papyri.

Transfer learning was used with GreekBERT, a Transformer that is pre-trained in masked language modelling, among other tasks, in modern Greek (Koutsikakis et al., 2020). GreekBERT has been further pre-trained in ancient Greek (Singh et al., 2021). We experimented with fine-tuning both variants in predicting the date,[8] but MAE was approx. one century (Table 2).

5.2 Regression analysis

Experiments were undertaken with Google Colaboratory, using a 12GB NVIDIA Tesla K80 GPU. We extracted term-frequency-inverse-document-frequency features using lower-cased text and character n-grams (from 1- to 5-grams).[9] All other parameters were set to default values.[10]

Monte Carlo cross validation

Linear regression achieved a MAE of 86 years on average (Table 2) and a MSE of 1.33. R^2 was similar across folds, around 83. A random forest had an even better MAE of 73 years on average but a worse MSE (1.58). Its average R^2 was lower than that of linear regression, but the maximum one achieved across folds was much better. Random forest also outperformed both gradient boosting methods in MAE but GBoost achieved a better MSE and R^2 on average. XTrees achieved the best results in all metrics, with a MAE of 54 years and the best R^2 climbing up to 95.43.

Leave one out cross validation

Using the best performing XTrees, we performed leave one out cross validation, by hiding one instance, training the algorithm on the remaining instances, and then using the model to predict the hidden record.[11] The MAE was found to be 55 years, MSE was 1.11, and R^2 was 85.89, close to the Monte Carlo evaluation scores. In order to better understand the errors, we rounded the predictions and the ground truth, evaluating as if we would in a classification setting. Predictions most often fall on or close to the diagonal (Figure 4), which explains the low error. The best result is achieved for the 1st and 2nd CE, followed by the 7th CE (see Table 3). The overall accuracy is 60%.

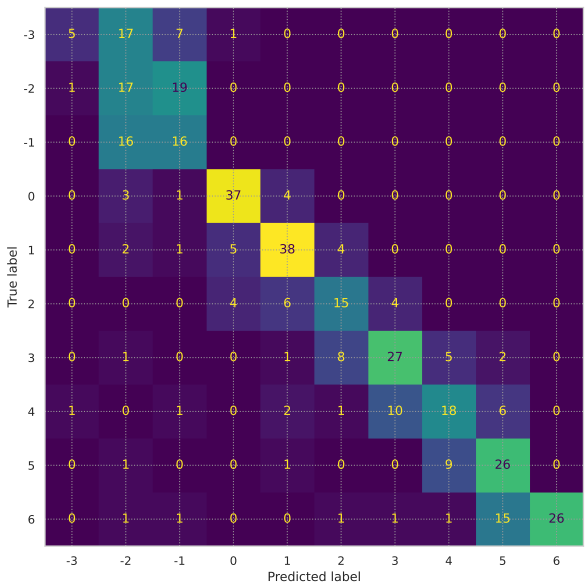

Figure 4: Confusion matrix of the extremely randomised trees (leave one out) when predictions and ground truth are rounded to century labels, see Table 3.

Error analysis

In very few cases, our leave-one-out regression fell considerably out of its predictions (Figure 4). Our analysis showed that these texts happen to contain specific words typical of another period, which confused the prediction. For instance among the highest prediction error were two late texts (6-7thCE) that exceptionally contain Σεραπίου and Βασιλείου, usually found in Ptolemaic time (3rd-1stBCE). In another case, we provided experimentally the longer version of the text, initially parsed only partially (§3.4). Using the full text led to an accurate prediction, influenced by the word 'indiction' in the additional text (§7.1).

6 Use cases

We applied our 389 regressors, produced upon leave-one-out cross-validation, to three use cases, which present different types of dating challenges.

6.1 PSI 8 934

This document[12] preserves the ca. 15 last lines of a land lease. The beginning of the text (the upper part of the sheet), where dating formulas are usually located, is thus missing. Nevertheless, the document can be securely attributed to a well-known group of

[8] We used white space, to allow subword computation.

[9] Preliminary experiments with centroid or trainable word embeddings before recurrent or convolutional neural networks deteriorated performance.

[10] Manual hyper-parameter tuning per regressor yielded insignificant improvements.

[11] The experiment lasted 15 hours.

[12] https://papyri.info/ddbdp/psi;8;934

	MAE↓		MSE↓		R^2 ↑	
	min	avg	min	avg	max	avg
Linear	0.73	0.86 (0.04)	0.92	1.33 (0.12)	85.34	82.72 (1.19)
Forest	0.65	0.73 (0.04)	0.93	1.58 (0.22)	89.53	79.12 (2.98)
GBoost	0.75	0.80 (0.02)	1.07	1.41 (0.12)	87.94	81.41 (1.99)
XGBoost	0.68	0.83 (0.06)	1.22	1.72 (0.23)	85.25	77.04 (3.40)
XTrees	**0.45**	**0.54 (0.03)**	**0.41**	**1.17 (0.26)**	**95.43**	**84.64 (4.22)**
Ithaca-max†	1.04		2.79		64.54	
Ithaca-avg†	0.97		2.33		69.98	
mGreekBERT†	1.11		1.91		76.59	
aGreekBERT†	0.91		2.03		75.17	
EfficientNet‡	2.05		7.73		8.75	
Vanilla CNN‡	3.66		20.92		-1.48	

Table 2: Minimum and average (st. error of the mean) MAE, MSE, and R^2 per regressor, best results per column in bold. The lower part of the table includes results from preliminary experiments with text (†) and image (‡) baselines.

Label	Century	Precision	Recall	F1
-3	3BCE	0.82	0.29	0.43
-2	2BCE	0.33	0.51	0.40
-1	1BCE	0.33	0.44	0.37
0	1CE	0.77	0.82	0.80
1	2CE	0.80	0.78	0.79
2	3CE	0.47	0.48	0.47
3	4CE	0.63	0.55	0.59
4	5CE	0.57	0.55	0.56
5	6CE	0.57	0.73	0.64
6	7CE	1.00	0.61	0.76

Table 3: Century-level classification evaluation.

texts from the 6th and early 7thCE c., the Dioscorus archive (Fournet, 2008), because, among other concordant elements, it contains microtoponyms from the respective village countryside. The notary who signed the contract, Abraam, is known from other documents, which is crucial evidence for the dating of the papyrus. This notary's period of activity has been proven to span at least between 524 and 545 (Fournet, 2003). This papyrus, therefore, is securely dated by indirect evidence, but no date is explicitly mentioned in the text (Fournet, 2008). Our average prediction is 310 CE, dated between 260 CE (min) and 352 CE (maximum prediction).

6.2 P. Basel 2 15

This papyrus, also shown in Figure 1, is a private letter dated indirectly from the 1st CE. The letter is almost complete, except for a damaged word at the end of line 5. Private letters usually do not bear a date. The dating, therefore, by the editor is done on palaeographical grounds as well as on the basis of scribal habits: "the hand [...] is more at home in the first century CE than the second, a dating that is supported by the writer's use of iota adscript..."(Huebner et al., 2020). Iota adscript is an expected feature in the 3rd BCE, starting to be irregularly written between the 2nd BCE and the first CE to almost completely disappear from the 2nd CE onwards (Clarysse, 1976). Onomastics strengthen the editor's dating hypothesis: of the three personal names mentioned in the letter (Pasis, Orsenouphis, and Tithoes), the first two are attested from ca. 250 BCE to 250 CE while the last one starts appearing in the papyri only in the 1st c. CE.[13] Our models date this to 140 BCE, from 165 BCE to 112 BCE.

6.3 P. Petra 1 5

The last manuscript[14] contains a request for transfer of taxation from 538 CE. It is a geographical outsider since it does not come from Egypt but from Petra (Jordan). We tested this manuscript since many of the words found in the text are infrequent in Egyptian manuscripts, on which our models are trained. The date mentioned in the papyrus is "second indiction". This refers to the second year of a repeated fifteen-year cycle (indiction) and the year 538 is relative, since it could be the second year of the previous or the next indiction (523 or 553). 538 is logically deduced by the editors in view of the whole dossier of papyri from Petra. Our models date this manuscript to 555 CE (521-575 CE),

[13] https://www.trismegistos.org/namvar/5135
[14] http://papyri.info/ddbdp/p.petra;1;5

overcoming the geographical variation.

7 Discussion

The computational, quantitative method suggested in this work is intended to complement human expertise. Its main contribution lies in providing an additional dating criterion for ancient Greek documents, in addition to the ones usually employed by papyrologists (palaeography, onomastics, prosopography, toponymy, archaeological evidence, etc.). It can predict a date for those papyri that do not include one, narrow down the possible time-span of doubtful dating, or contribute to deciding on one particular date when several alternatives seem possible. Despite the fact that limitations exist (discussed in §7.3), compared to traditional approaches the models trained in this study are expected to reduce biases. Their value is not limited to predicting dates for individual manuscripts, but they can be applied to any attribute of a group of papyri, e.g. the place of provenance or the text's type. At the same time, easily accessible open-source metadata exist for most published papyri (§3.1).

7.1 Rationale generation

The use of supervised learning, such as the work of Assael et al. (2022) or ours, can yield accurate estimations, which can at least help the human expert. The assistance is greater, however, when explanations are provided for the models' decisions. In our case, we used a committee of hundreds of regressors in order to estimate the date of three use cases. Therefore, we sampled models per case and generated rationales regarding their predictions, by using their Shapley values (Lundberg and Lee, 2017). In the case of PSI 8 934 (§6.1), our investigation showed that the mention of the name 'Aurelios Victor' ('Αὐρήλιος Βίκτωρ') influenced the decision, resulting to a more recent date than what would have been predicted otherwise. Similarly, in the case of P. Petra 1 5 (§6.3), the decision was influenced by a reference to 'indiction' ('ἰνδικτίωνος'), a word that refers to a periodic reassessment of taxation in the Late Roman Empire.

7.2 In the wild

Computational dating can facilitate a macroscopic analysis of vaguely dated or undated manuscripts. By generating estimated dates for hundreds of such manuscripts, the expert can view from distance the collection, potentially drawing useful conclusions or making significant remarks. To test this hypothesis, we collected 220 manuscripts dated with an upper CE date limit (i.e., not after that date). We formed a committee of regressors,[15] and we estimated the minimum, the maximum, and the average chronology of each manuscript. In 28% of them, the maximum prediction exceeded the upper threshold and was discarded to avoid doubting the expert. This process led to the date estimation for 159 manuscripts, which we release publicly in our repository to assist other researchers. As can be seen in Figure 5, some of our estimations fall far away from the upper limit (in red) while others fall close. The estimated date from our regressors' committee should be read along with other information, which is kept in the shared corpus, such as the place settlement (Figure 6 shows frequent places). We observe, for example, that in some places the estimated dates fall closer to the upper limit (e.g. in Oxyrhynchos and Tebtynis the distance is 132 years) compared to others (e.g. in Antinoopolis and Hermopolis the distance is 283 and 384 years).

7.3 Challenges and limitations

Our experimental analysis proved that text regression is a considerably reliable and accurate tool in dating nonliterary papyri. Limitations and challenges stem mainly from the composition of our dataset, which is balanced as far as the dates of the papyri included are concerned, both at the level of the century (approx. 40 records per century) and at the level of the quarter of the century (albeit less strictly and with the exception of the 7th CE). Furthermore, although we retained a substantial text sample of each papyrus, in approximately 1/4 of the records some text was eliminated.

Biases

Despite our effort to balance the dataset in terms of dates, biases are present. Since our main concern in collecting the data was for the date distribution, no deliberate selection was made on the basis of the document types. Some types are thus over or underrepresented (e.g. private letters that do not usually bear a date; §6.2). Each type of document has however distinctive linguistic characteristics, such as the level of formality or unusual constructions (e.g. accounts). This uneven typological representation probably affects the performance of the models. Other possible biases in the dataset concern the

[15]We sampled randomly 100 regressors.

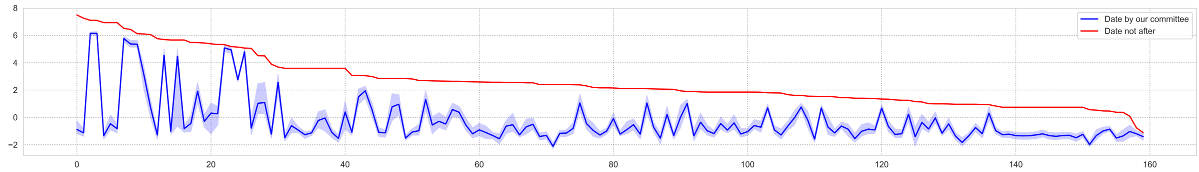

Figure 5: Date estimations by a committee of regressors, with minimum and maximum shadowed. In red is the upper limit for the date, which was already known for these manuscripts.

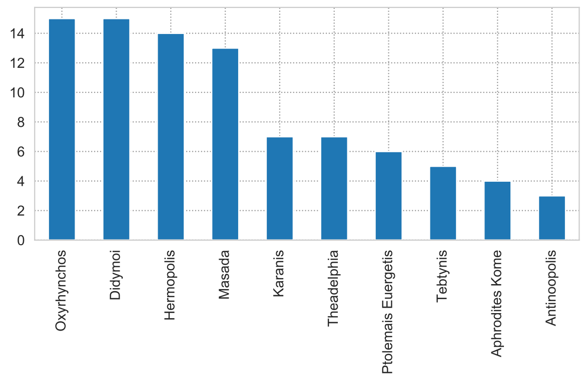

Figure 6: The frequency of settlement places in the corpus that was dated by our committee of regressors.

provenance of papyri, the length of their text, and the state of conservation (sizeable portions of missing text or entirely missing parts of the documents).

Chronological analysis of words

Chronological analysis of word occurrence is possible if we detect and collect terms only attested in the papyrological material during a limited period. The word 'denarius' only appears after the 2nd CE and before the 5th CE, its presence in a text thus means that the text must have been written during this timespan. Likewise a text containing the word 'indiction' cannot have been written before the 4th CE. The investigation should also regard the possibility that the models make a prediction for a papyrus based on typical dating formulas present in the text like the name of the ruling emperor. Although our investigation of explanations did not yield any major concerns, a bigger sample of test cases should be created and more explainability methods should be employed (Ribeiro et al., 2016) to make conclusive remarks on this front.

Transcription of papyri is not optional

Transcription of the papyri is required (at least partial, but substantial) to reach this high degree of accuracy with our method. Thus, while there are transcriptions available for most already published papyri, it is less practical for dating unpublished papyri that have not been yet transcribed to a relatively high standard. In that case, image classification on the scripts can provide a less accurate prediction of the date as starting point.

8 Conclusion

We presented a machine-actionable dataset of 389 Greek documentary papyri of (mostly) Egyptian provenance, dated and balanced in terms of chronological quarter-century distribution. We trained extremely randomised trees on top of character n-gram-based features, reaching a mean absolute error of 54 years and 60% in century-level classification accuracy. We then formed a committee of regressors, which we applied to three use cases: a land lease, a private letter, and a geographical outsider (not from Egypt). To assist future research, our committee dated 159 manuscripts, for which only the upper limit is known. Future endeavours for this research extend far beyond the dating of individual manuscripts. It can produce valuable data for the study of the Greek language and its evolution through a millennium, help identify and trace linguistic habits and trends, as well as the history of document production, circulation, and use (e.g. which period produces what kind of texts, which administration relied on what type of documents, etc.). It can also produce further data and resources towards the typology of ancient Greek documents, completing with computational methods the work already underway and well-advanced of the grammateus project. Last, it can in the future fruitfully be combined with computational paleography to analyse the script and content of a given text.

Acknowledgements

This work was partially supported by the Swiss National Science Foundation as part of the project no. PZ00P1-174149 "Reuniting fragments, identifying scribes and characterising scripts: the Digital paleography of Greek and Coptic papyri (d-scribes)"

References

Kalthoum Adam, Asim Baig, Somaya Al-Maadeed, Ahmed Bouridane, and Sherine El-Menshawy. 2018. Kertas: dataset for automatic dating of ancient arabic manuscripts. *International Journal on Document Analysis and Recognition (IJDAR)*, 21:283–290.

Yannis Assael, Thea Sommerschield, Brendan Shillingford, Mahyar Bordbar, John Pavlopoulos, Marita Chatzipanagiotou, Ion Androutsopoulos, Jonathan Prag, and Nando de Freitas. 2022. Restoring and attributing ancient texts using deep neural networks. *Nature*, 603(7900):280–283.

Anaëlle Baledent, Nicolas Hiebel, and Gaël Lejeune. 2020. Dating ancient texts: an approach for noisy french documents. In *Language Resources and Evaluation Conference (LREC) 2020*.

Malcolm Choat. 2019. Dating papyri: Familiarity, instinct and guesswork1. *Journal for the Study of the New Testament*, 42(1):58–83.

Willy Clarysse. 1976. Notes on the use of the iota adscript in the third century bc. *Chronique d'Egypte*, 51(101):150–151.

Maruf A Dhali, Camilo Nathan Jansen, Jan Willem De Wit, and Lambert Schomaker. 2020. Feature-extraction methods for historical manuscript dating based on writing style development. *Pattern Recognition Letters*, 131:413–420.

Manuel Fernández-Delgado, Manisha Sanjay Sirsat, Eva Cernadas, Sadi Alawadi, Senén Barro, and Manuel Febrero-Bande. 2019. An extensive experimental survey of regression methods. *Neural Networks*, 111:11–34.

Jean-Luc Fournet. 2003. P. Köln Gr. X 421. *Kölner Papyri*, (Band 10).

Jean-Luc Fournet. 2008. *Les archives de Dioscore d'Aphrodité cent ans après leur découverte. Histoire et culture dans l'Égypte byzantine*. De Boccard.

Pierre Geurts, Damien Ernst, and Louis Wehenkel. 2006. Extremely randomized trees. *Machine learning*, 63(1):3–42.

Anmol Hamid, Maryam Bibi, Momina Moetesum, and Imran Siddiqi. 2019. Deep learning based approach for historical manuscript dating. In *2019 International Conference on Document Analysis and Recognition (ICDAR)*, pages 967–972. IEEE.

Anmol Hamid, Maryam Bibi, Imran Siddiqi, and Momina Moetesum. 2018. Historical manuscript dating using textural measures. In *2018 International Conference on Frontiers of Information Technology (FIT)*, pages 235–240. IEEE.

Hermann Harrauer. 2010. *Handbuch der griechischen Paläographie*. Hiersemann.

Sheng He, Petros Samara, Jan Burgers, and Lambert Schomaker. 2016a. Historical manuscript dating based on temporal pattern codebook. *Computer Vision and Image Understanding*, 152:167–175.

Sheng He, Petros Samara, Jan Burgers, and Lambert Schomaker. 2016b. Image-based historical manuscript dating using contour and stroke fragments. *Pattern Recognition*, 58:159–171.

Sheng He, Petros Sammara, Jan Burgers, and Lambert Schomaker. 2014. Towards style-based dating of historical documents. In *2014 14th International Conference on Frontiers in Handwriting Recognition*, pages 265–270. IEEE.

Sabine R Huebner, W Graham Claytor, Isabelle Marthot-Santaniello, and Matthias Müller. 2020. *Papyri of the University Library of Basel (P. Bas. II)*, volume 41. Walter de Gruyter GmbH & Co KG.

John Koutsikakis, Ilias Chalkidis, Prodromos Malakasiotis, and Ion Androutsopoulos. 2020. Greek-bert: The greeks visiting sesame street. In *11th Hellenic Conference on Artificial Intelligence*, pages 110–117.

Yuanpeng Li, Dmitriy Genzel, Yasuhisa Fujii, and Ashok C Popat. 2015. Publication date estimation for printed historical documents using convolutional neural networks. In *Proceedings of the 3rd international workshop on historical document imaging and processing*, pages 99–106.

Scott M Lundberg and Su-In Lee. 2017. A unified approach to interpreting model predictions. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems 30*, pages 4765–4774. Curran Associates, Inc.

Roberta Mazza. 2019. Dating early christian papyri: Old and new methods–introduction. *Journal for the Study of the New Testament*, 42(1):46–57.

George Mikros, Nick Hatzigeorgiu, and George Carayannis. 2005. Basic quantitative characteristics of the modern greek language using the hellenic national corpus. *Journal of Quantitative Linguistics*, 12(2-3):177–177.

Brent Nongbri. 2014. The limits of palaeographic dating of literary papyri: Some observations on the date and provenance of p.bodmer ii (p66). *Museum Helveticum*, 71(1):1–35.

Brent Nongbri. 2019. Palaeographic analysis of codices from the early christian period: A point of method. *Journal for the Study of the New Testament*, 42(1):84–97.

Enock Osoro Omayio, Sreedevi Indu, and Jeebananda Panda. 2022. Historical manuscript dating: traditional and current trends. *Multimedia Tools and Applications*, pages 1–30.

Asimina Paparigopoulou, John Pavlopoulos, and Maria Konstantinidou. 2022. Dating greek papyri images with machine learning, 17 november 2022, preprint (version 1). *Research Square*.

Iiro Rastas, Yann Ciarán Ryan, Iiro Tiihonen, Mohammadreza Qaraei, Liina Repo, Rohit Babbar, Eetu Mäkelä, Mikko Tolonen, and Filip Ginter. 2022. Explainable publication year prediction of eighteenth century texts with the BERT model. In *Proceedings of the 3rd Workshop on Computational Approaches to Historical Language Change*, pages 68–77, Dublin, Ireland. Association for Computational Linguistics.

Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. 2016. " why should i trust you?" explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 1135–1144.

Pranaydeep Singh, Gorik Rutten, and Els Lefever. 2021. A pilot study for bert language modelling and morphological analysis for ancient and medieval greek. In *The 5th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature (LaTeCH-CLfL 2021)*.

Mingxing Tan and Quoc Le. 2019. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning*, pages 6105–6114. PMLR.

Fredrik Wahlberg, Lasse Mårtensson, and Anders Brun. 2015. Large scale style based dating of medieval manuscripts. In *Proceedings of the 3rd International Workshop on Historical Document Imaging and Processing*, pages 107–114.

Fredrik Wahlberg, Tomas Wilkinson, and Anders Brun. 2016. Historical manuscript production date estimation using deep convolutional neural networks. In *2016 15th International Conference on Frontiers in Handwriting Recognition (ICFHR)*, pages 205–210. IEEE.

ACL 2023 Responsible NLP Checklist

A For every submission:

☑ A1. Did you describe the limitations of your work?
7.3

☑ A2. Did you discuss any potential risks of your work?
7.3

☑ A3. Do the abstract and introduction summarize the paper's main claims?
1

☒ A4. Have you used AI writing assistants when working on this paper?
Left blank.

B ☑ Did you use or create scientific artifacts?

3

☑ B1. Did you cite the creators of artifacts you used?
3

☑ B2. Did you discuss the license or terms for use and / or distribution of any artifacts?
3

☑ B3. Did you discuss if your use of existing artifact(s) was consistent with their intended use, provided that it was specified? For the artifacts you create, do you specify intended use and whether that is compatible with the original access conditions (in particular, derivatives of data accessed for research purposes should not be used outside of research contexts)?
7

☐ B4. Did you discuss the steps taken to check whether the data that was collected / used contains any information that names or uniquely identifies individual people or offensive content, and the steps taken to protect / anonymize it?
Not applicable. Left blank.

☑ B5. Did you provide documentation of the artifacts, e.g., coverage of domains, languages, and linguistic phenomena, demographic groups represented, etc.?
3

☑ B6. Did you report relevant statistics like the number of examples, details of train / test / dev splits, etc. for the data that you used / created? Even for commonly-used benchmark datasets, include the number of examples in train / validation / test splits, as these provide necessary context for a reader to understand experimental results. For example, small differences in accuracy on large test sets may be significant, while on small test sets they may not be.
3

C ☑ Did you run computational experiments?

5

☑ C1. Did you report the number of parameters in the models used, the total computational budget (e.g., GPU hours), and computing infrastructure used?
5

The Responsible NLP Checklist used at ACL 2023 is adopted from NAACL 2022, with the addition of a question on AI writing assistance.

☑ C2. Did you discuss the experimental setup, including hyperparameter search and best-found hyperparameter values?
5

☑ C3. Did you report descriptive statistics about your results (e.g., error bars around results, summary statistics from sets of experiments), and is it transparent whether you are reporting the max, mean, etc. or just a single run?
Sections 5 and 7

☑ C4. If you used existing packages (e.g., for preprocessing, for normalization, or for evaluation), did you report the implementation, model, and parameter settings used (e.g., NLTK, Spacy, ROUGE, etc.)?
4

D ☒ Did you use human annotators (e.g., crowdworkers) or research with human participants?

Left blank.

☐ D1. Did you report the full text of instructions given to participants, including e.g., screenshots, disclaimers of any risks to participants or annotators, etc.?
No response.

☐ D2. Did you report information about how you recruited (e.g., crowdsourcing platform, students) and paid participants, and discuss if such payment is adequate given the participants' demographic (e.g., country of residence)?
No response.

☐ D3. Did you discuss whether and how consent was obtained from people whose data you're using/curating? For example, if you collected data via crowdsourcing, did your instructions to crowdworkers explain how the data would be used?
No response.

☐ D4. Was the data collection protocol approved (or determined exempt) by an ethics review board?
No response.

☐ D5. Did you report the basic demographic and geographic characteristics of the annotator population that is the source of the data?
No response.