"Alpha, Aleph and AI: Languages of the Ancient Mediterranean and Near East" Conference

Dates: Wed and Thurs 14th and 15th June 2023
Location: Arts Complex, 7 Woodland Road, University of Bristol, BS8 1TB

Registration free: in-person and/or zoom via Eventbrite

ABSTRACTS OF ALL TALKS (in alphabetical order of speaker[s])

Tsehay Ademe Belay (Addis Ababa University)
"Comparative Semitics philological inquiry based on the Ethiopic Book of Joel"

The translation of the Bible into classical Ethiopic is a landmark in Ethiopian spiritual literature. In order to notice the remarkable appreciation of biblical prophets in Ethiopia; in the Ethiopic manuscript tradition, the Book of Joel is one of the books of the 12 Minor Prophets following the Book of Daniel. The Book of Joel is famous for its call to repentance and correction of interior life during fasting time symbolising the Day of the Lord through locust swarms, battles, and powerful imagery. If we take Joel 1:15, we will observe significant differences between several Ethiopic manuscripts. There have been words and phrases which are not existed in Hebrew, Syriac/Aramaic and Arabic. Behind UNESCO10_34, one obtains a reading which expresses pain and the imminence of some danger without precision as to who is suffering. On the other hand, EMML1768 affirms that it is the prophet who suffers. As for IES0077 and the warning is directly addressed to the people, without including the prophet. Now, how is it possible to get such differences if the translation came from one text or one original language? Or can one explain these differences by taking recourse to different source texts and languages? On the one hand, a critical edition will enable us to identify different families and to trace the textual history of the Ethiopic Book of Joel. In this connection, this paper intends to show the significance elements of word innovations in the Ethiopic Book of Joel.

Riccardo Bongiovanni (University of Pisa)
"Tools for a new digital edition of the *Corpus of Iatromagical Papyri*"

In my paper, I will showcase the opportunities offered by the READ software, a digital tool to study ancient documents; in particular, I will focus on papyri. READ (Research Environment for Ancient Documents) is a multifunctional software developed by Stefan Baums, Andrew Glass and Stephen White which elaborates data taken from physical evidence such as papyri, codices, ostraka. The main function of the software is to catalogue ancient documents and to categorise them using 'tags' (e.g. Default; BaseType; VowelType; FootmarkType). Once catalogued, the texts constitute a database to draw from to, to put it simply, identify scripts. Single letters extrapolated from different documents can thus be easily compared to offer palaeographic insight. It is also possible to personalise the database through the

aforementioned tags. Through READ it is possible to have several transcription layers open simultaneously, and to sift through different interpretations and reconstructions of a text at the same time.

In my paper, I will show practical examples of the application of this tool using the *Greek Iatromagical Papyri,* a subset of the *Greek Magical Papyri*, which combine the ritual and religious vocabulary and layouts of magical and scientific texts. Using READ I have been able to develop a digital edition in which, thanks to the tags, each word of the transcription was immediately identifiable in the papyrus, even with multiple textual reconstruction options. I'm also creating a classification of the *charakteres*, magical symbols with a still uncertain meaning, whose interpretation is an important part of my current PhD project.

Ivri J. Bunis (University of Haifa)

"Parallel Morphophonemic Consequences of Guttural Weakening in Hebrew and Western Neo-Aramaic"

My talk will present a comparative-historical analysis of guttural weakening in Hebrew and in Western Neo-Aramaic. Specifically, I shall compare the effects of guttural weakening on certain morphological categories in both languages, especially in combinations of a proclitic preposition + guttural initial word. Guttural consonants or laryngeals are documented to have been relatively weak in Hebrew, especially in its Late Biblical and post-Biblical periods and in various dialects of Aramaic. The weakening of the gutturals takes various forms, especially elision, but also their merger with other consonants or inability to be geminated. In Hebrew, in its Masoretic Biblical Hebrew traditions and Mishnaic Hebrew traditions, and in Western Neo-Aramaic, guttural weakening is not absolute. Guttural consonants show certain signs of weakening yet are largely retained. In both languages, it is mainly the glottal consonants /h/ and /ʔ/ that show weakness, in specific phonological environs. My talk will aim to demonstrate two main points. Firstly, the consonants /h/ and /ʔ/ underwent similar sound changes in Hebrew and in Western Neo-Aramaic. Secondly, in both languages, those parallel sound changes brought about similar morphological and morphophonemic developments. I propose that these Hebrew-Western Neo-Aramaic commonalities illuminate the general linguistic reality of Syro-Palestine at the beginning of the common era. This proposal has historic justification, namely, around the beginning of the common era, Hebrew was in close contact with Western Aramaic dialects from Syro-Palestine. The closest existing representative of Western Aramaic is Western Neo-Aramaic.

Minqi Chu (Sorbonne Université)

"Editing, translation, and commentary on a glossary of anatomy in the Italo-Greek manuscript Paris. gr. 1053"

The Italo-Greek manuscript Paris. gr. 1053, containing many ascetic-dogmatic texts, was transcribed at a Greek monastery founded by S. Neilos the Young in Campania in the 10th-11th century, and in the blank areas of f. 106v and f. 138r, another hand of a slightly later date transcribed two passages entitled *Ἐτυμολογία ἀπό κεφαλὴς ἕως ποδῶν* and *Ἐτυμολογία τῶν ἐντὸς τοῦ σώματος*, which formed a minor etymological glossary of the human body. It shows a

number of etymological interpretations that completely differ from other Greek *Etymologica* in the ancient and Byzantine world. This paper will first edit this short etymological glossary, then examine and comment on its etymologies through comparisons with other ancient and Byzantine etymological lexicons, and finally place its contents in the context of ancient Greek works on medicine and physiology for further analysis.

Robert Crellin (Classics, Cambridge)

"How important is text type in training models for morphosyntactic annotation of epigraphic texts?"

In light of recent advances in the treebanking of Greek and Latin texts (Haug & Jøhndal 2008; Celano, Crane & Almas 2014; Keersmaekers et al. 2019; Gorman 2020; Vierros & Henriksson 2021; Pitts 2022), I set out the methodology and interim results of work-in-progress morphosyntactically annotating Greek and Latin texts from *I.Sicily* (Prag 2022). I discuss the extent to which models trained on non-epigraphic texts are helpful for the automated analysis of epigraphic texts. The issue is pertinent since research involving treebanks on ancient languages is often poorly resourced compared with modern language equivalents (see Johnson et al. 2021: 21).

My annotation process itself is iterative. First a subset of the corpus is annotated manually. A model is built on the basis of this subset, which is then used to annotate another subset. Once this second subset is corrected, a new model is built. The process is to be repeated until the entire corpus is annotated.

An important benchmark for measuring model effectiveness is the CLTK parser (Johnson et al. 2021) which is trained on literary texts (for corpora see https://github.com/cltk). My *ad hoc* attempts to-date to annotate epigraphic texts using CLTK have not proved successful, suggesting that similarity of text type may be a problem. To see if this is indeed the case, I compare the results obtained by applying the iterative procedure described above with those obtained by applying CLTK.

References

Celano, G. G. A., Crane, G. and Almas, B. (eds) (2014) *The Ancient Greek and Latin Dependency Treebank*. https://perseusdl.github.io/treebank_data/

Gorman, V. B. (2020) 'Dependency treebanks of Ancient Greek prose', *Journal of Open Humanities Data* 6, 1–3. doi: 10.5334/johd.13

Haug, D. T. T. and Jøhndal, M. L. (2008) "Creating a parallel treebank of the old Indo-European Bible translations." In Sporleder, C. and Ribarov, K. (eds) *Proceedings of the Second Workshop on Language Technology for Cultural Heritage Data* (LaTeCH 2008), 27–34.

Johnson, K. P., Burns, P. J., Stewart, J., Cook, T., Besnier, C. and Mattingly, W. J. B. (2021) 'The Classical Language Toolkit: An NLP Framework for Pre-Modern Languages.' In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing: System Demonstrations. Online: Association for Computational Linguistics*, 20–29. doi: 10.18653/v1/2021.acl-demo.3

Keersmaekers, A., Mercelis, W., Swaelens, C. and Van Hal, T. (2019) 'Creating, enriching and valorising treebanks of Ancient Greek: The ongoing Pedalion-project.' In *Proceedings of the 18th International Workshop on Treebanks and Linguistic Theories* (TLT, SyntaxFest 2019). Paris: Association for Computational Linguistics, 109–117. doi: 10.18653/v1/W19-7812

Pitts, R. (2022) 'Corpus of the Epigraphy of the Italian Peninsula in the 1st Millennium BCE (CEIPoM)', *Journal of Open Humanities Data* 8(1), 1–4. doi: 10.5334/johd.65

Prag, J. R. W. (ed.) (2022) *I.Sicily*: . http://sicily.classics.ox.ac.uk; doi: 10.5281/zenodo.4021517

Vierros, M. and Henriksson, E. (2021) 'PapyGreek treebanks: A dataset of linguistically annotated Greek documentary papyri', *Journal of Open Humanities Data* 7, 1–6. doi: 10.5334/johd.55

Mark Depauw (KU Leuven) (KEYNOTE)
"Trismegistos: facilitating (quantitative) research across languages of the Ancient Western World"

I will present the Trismegistos project: its origins, goals, history and current developments, with particular attention for multilingualism.

Luís Firmino (University of São Paulo, Brazil)
"The Article with Proper Names in Herodotus and Thucydides"

In this presentation, we compare the use of article with proper names in the works of Herodotus and Thucydides, by selecting five names in each author (which amount to 920 tokens in both) and assessing the reference to those names, through the analysis of the context in which they appear. Recent works that dealt with Definiteness in Ancient Greek have contributed to the understanding of the presence or absence of article in AG, but the use with proper names requires further examination. To our knowledge, only Rijksbaron has dedicated a study to this matter, which carries valuable contributions, though still limited. Here, we examine the presence or absence of the article in relation to a number of linguistic features present in the context of appearance of the proper name (such as the case of the name, its agentivity, whether it commands a speech verb, among others). Our goal is to show, through a quantitative examination of the data, added to the qualitative analysis of some examples, that some linguistic (Semantic, Pragmatic, Discursive) features can be closely related to the presence of the article with proper names. Furthermore, we suggest that the difference in the use of the article by each author and with each name can be seen as a tool for characterization. The ongoing research in which this presentation is based is funded by São Paulo State Research Foundation (FAPESP), grant number 2021/15096-0.

Stefan Hagel (Austrian Academy of Sciences) (KEYNOTE)
"Digitally Editing Ancient Music"

An edition of genuine musical sources such as the fragments of notation from Classical Antiquity calls for an online, multimedia, potentially interactive, approach like little else. I will

focus on the challenges and potential solutions of such a project, as well as its impact on ancient-music research.

Toby Hudson (Wolfson, Oxford)
"Using acoustic simulations to model stress shift in Latin"

It is understood that Old Latin, like the surrounding Italic languages and Etruscan, had a fixed word-initial stress in place of the inherited Indo-European accent. This is evidenced by word-medial syncope in reconstructed forms and in borrowed words such as *balneum* from βαλανεῖον. However, by the first century BC the 'penultimate' stress system, familiar from the Romance languages, had emerged, in which primary word stress fell on the antepenultimate syllable unless the following syllable was metrically heavy. What caused this wholesale reconfiguration of word prosody such as *fárci:re* to *farcí:re*? Possible explanations include the swapping of primary and secondary stress, and the metrical re-parsing of words. Here, however, I seek phonetic support for the notion that a right-edge stress system might emerge from micro-fluctuations in the production of stress.

There are two key elements to this investigation. First, using prosodic minimal pairs in Italian as a proxy (e.g. pássero 'sparrow', passeró 'I shall pass') I interpolate between these using a vocoding in Mel-Frequency Cepstral Coefficients with a technique to generate natural-sounding audio files from a continuum of vectors which fall between the two inputs. The second is to play the files generated in this way to listeners – native speakers of Italian – in randomised sequence, with a binary choice as to which word they hear. I present evidence of a categorical boundary in the judgments which supports the possibility of categorical change in word stress emerging from the perception of small variations in the parameters of accentuation.

Aaron Koller (Yeshiva University)
"Alphabetical Order and Alphabetical Thinking from Antiquity to the Middle Ages: The History of an Idea and the Paths of Transmission"

Can we write the history of alphabetical order? The alphabet probably had an order already at the moment of its invention, roughly 4000 years ago, and evidence from the Levant suggests that it had two orders, which co-existed for nearly 2000 years. More important is the use of the alphabet meta-linguistically, not just as signs of sounds but as a principle of organization. This is found in acrostic poems of the Iron Age and later. Evidence from Alexandrian papyri and Demotic ostraca shows that lists, too, could be alphabetically arranged by Hellenistic and Roman times.

In Late Antiquity, the Latin West more or less ignored the metalinguistic uses of the alphabet, perhaps because the Latin alphabet in use was in any event not believed to be the original one. In the east, however, the alphabet was intensively studied and contemplated; Samaritan, Syriac Christian, Jewish, Mandaic, Byzantine Greek, and Coptic texts all reflect on the shape and order of the alphabet. It was utilized practically, as well, as in Arabic lexicography from its very beginnings in the eighth century and through Muslim, Jewish, and Byzantine Greek dictionaries of the Middle Ages.

Alphabetic thinking returns to western Christendom in the twelfth century. There is ample reason to think that alphabetization, too, arrived in Europe around 1200 from the Semitic world. The Arabic tradition of alphabetization is likely to be the source of the rise in Latin alphabetization visible to medievalists beginning in the thirteenth century. This was probably mediated through Jewish books, which traveled, physically and intellectually, from the Islamicate world to that of Latin Christendom, bringing many of their metatextual practices along with them. The vibrant book cultures of the Jews and Muslims left a deep impression on the Christian book culture of the High Middle Ages.

Maria Konstantinidou (Democritus University of Thrace)

"A preliminary authorship analysis of the New Testament"

This paper presents the initial results of an exploratory study of the New Testament, using statistical language models for authorship attribution and analysis. Computational methods for authorship questions have been applied recently to Greek texts (for example for determining the author of the pseudo-Euripidean play *Rhesus*"[1] or to address the Homeric Question[2]).

In this study, the 27 books of the New Testament are put to the test with statistical language models trained on character trigrams and computing PPL (perplexity) of the passages/works under evaluation. A number of experiments are conducted, aiming at confirming that each book (or group of books) exhibits a distinct linguistic identity. They are then compared against other books attributed to the same author (for instance the 13 Pauline Epistles, of which only 7 are considered genuine). In another set of experiments, books attributed to different authors but considered to be linguistically similar (such as Luke's Gospel and the Acts of the Apostles) are also compared.

References

[1] N. Manousakis, E. Stamatatos, Devising Rhesus: A strange 'collaboration' between Aeschylus and Euripides, *Digital Scholarship in the Humanities*, Volume 33, Issue 2, June 2018, 347–361, https://doi.org/10.1093/llc/fqx021

[2] C. Bozzone, R. Sandell, One or many Homers? Using quantitative authorship analysis to study the Homeric Question, D. M. Goldstein, S. W. Jamison, Brent Vine (eds.), *32nd Annual UCLA Indo-European Conference*, Buske 2022, 21-48, https://doi.org/10.46771/978-3-96769-308-9 and J. Pavlopoulos, M., Konstantinidou, Computational authorship analysis of the Homeric poems, *International Journal of Digital Humanities*, 2022, https://doi.org/10.1007/s42803-022-00046-7

Todd Krause (University of Texas, Austin)

"Semitilex: One database to rule them all"

Group: Todd Krause, Na'ama Pat-El, and Danny Law (University of Texas, Austin)

The emergence of Digital Humanities has spurred a wealth of online tools and databases. However, detailed and accessible linguistic information on the Semitic family has

long been a desideratum for both the broader scholarly community and the general public. Despite continued interest, true general access to early Semitic languages remains limited.

To address both these scholarly and pedagogical lacunae, we are developing Semitilex. Semitilex is a comparative Semitic lexicon and annotated text collection, an accessible database with rich supporting annotations, glosses, and commentary. This resource will allow linguists, Semitic scholars, and the general public to easily incorporate insights about grammar and vocabulary, as well as contextualized readings of key primary texts in Semitic languages, into their research and other pursuits. The system also handles complex scripts; combines right-to-left writing alternating with left-to-right; and provides linguistic, cultural, and historical annotation appropriate for Semitists, linguists, and general audiences.

Semitilex incorporates traditional features (semantics, etymology), but expands them to include additional grammatical information: e.g. morphological patterns, gender, number morphology, verbal stem, thematic vowel, and types of verbal complements. This additional information increases the scholarly utility of the database, but also provides essential details for learners of Semitic languages, because the pattern a particular word uses to mark grammatical information has dramatic implications for what that word will mean and look like in a text.

In this presentation we will explain the rationale for the project, present its current state, and discuss future goals and benefits.

Orly Lewis and Premshay Hermon (Hebrew University of Jerusalem)
“Re-animating Greco-Roman Anatomy through Machine Learning and 3D Modelling”
Group: ATLOMY

The talk will discuss the application of data-science, NLP and software development in the study of Greco-Roman anatomical texts. Project ATLOMY is developing a web-based, open-access interactive platform which offers an integrative visual and textual map of Greco-Roman anatomical ideas and terminology. Our approach involves a data-centric analysis of ancient anatomical texts (in Greek, Latin and Arabic) using principles of NLP and data science. We identify anatomical words (nouns, verbs, prepositions, adjectives and more), extract the linguistic and anatomical features of each word, consider their textual, syntactical and semantic contexts and map them in relational tables. We do this both manually and with automated computational methods involving machine-learning models we have adapted to the languages and technical content of our sources.

Based on our lexical, textual and anatomical analysis we create 3D models of the ancient anatomical ideas described in the texts. These visual models are connected to our tokenised and mapped data, yielding a data-driven visualisation of the ancient anatomical ideas described in the texts. The output is made accessible and transparent through an interactive GUI tailored to researchers’ needs: they can search for any term, learn about its changing meanings in the history of medicine and philosophy, view references in various forms; sort and filter the search results by author, text or period; they can search and view 3D visualisations of the ancient anatomical ideas; swivel the models to observe it from all directions, remove parts to observe deeper layers; and investigate the names of the parts.

The talk will present our data-driven workflow from text to 3D model, with a focus on the data science and machine-learning components, demonstrate our interactive platform and discuss the reciprocal relation between its development and the textual research.

Lorenzo Livorsi (University of Bamberg)
"Law and Cursor: Digital Analysis of Prose Rhythm and Textual Criticism in the Late Roman Constitutions"

Rhythm played an important role in ancient literary prose. Recognisable patterns of long and short syllables (*clausulae*) marked clause endings, constituted a form of punctuation, and created compelling rhetorical effects. Over time, the phonetic/phonemic distinction between quantities was lost, and quantitative *clauisulae* were replaced by patterns of accented syllables. In the Middle Ages, these accentual endings were eventually formalised as *cursus*. But in Late Antiquity, the old and the new coexisted: rhythmic patterns tended to make the ancient quantitative *clausulae* coincide with the new accentual endings (*cursus mixtus*). As the most solemn genres, panegyrics and constitutions (i.e., laws composed for oratorical delivery), adhere systematically to prose rhythm.

Cursor, a new tool for the analysis of Latin prose (currently in beta testing), considers simultaneously both quantitative and accentual rhythm, and enables extended corpus comparisons. In my experience, I apply *Cursor* to late antique constitutions, including their abridgements in the Theodosian and Justinian Codes. Mapping prose rhythm in this corpus affects textual criticism. Mommsen and Meyer's otherwise unimpeachable edition fails to consider the prose rhythm. Yet the constitutions conform scrupulously to prose rhythm; hence – all else being equal – variant readings which form standard *clausulae* ought to be preferred. The recourse to prose rhythm as a discerning factor in establishing the text is not new, but the possibilities afforded by *Cursor* enable a fresh appraisal of methodological questions, including the role of prose rhythm in parsing the meaning, determining emphasis, dividing sentences correctly, spotting lacunae, and making conjectural criticism based on cursus.

Nikos Manousakis (Academy of Athens)
"Applying (automated) Authorship Attribution to Greek drama: An expanding variety of case studies"

In the last five years, I have applied a series of Authorship Attribution methods to disputed or partly disputed texts in the corpus of Greek drama: *Rhesus* attributed to Euripides, *Prometheus Bound* attributed to Aeschylus, the problematic ending of Aeschylus' *Seven against Thebes*, the controversial recognition scene in Euripides *Electra*, and, most recently, the fragments of *Prometheus Unbound*. The Authorship Attribution methods I have applied to these texts range from visualisation models (e.g., *Principal Components Analysis*, *Agglomerative Hierarchical Clustering*) to Artificial Intelligence algorithms (e.g., *Support Vector Machines*, the *Common n-grams* classification technique). In this paper I will discuss my published and unpublished work on the subject: that is, the trends, the promises, the challenges, and the gains

in employing this approach, and also some, quite unexpected, collateral benefits of this type of quantification of texts.

Willem Th. van Peursen (Vrije Universiteit Amsterdam) (KEYNOTE)
"Back to the Black Box? Explainable and Unexplainable AI in Semitic Studies"

This presentation will give a bird's eye view on the development of the field of "Bible and Computer" as it emerged in the 1970s till the most recent attempts to apply AI techniques to Biblical Hebrew and other Semitic languages. It will highly the paradox of the black box: The pioneers of the computer-aided textual analysis of the Hebrew Bible aimed at transparency and traceability. Following scholars such as James Barr and Jaap Hoftijzer, they strived for a data-driven approach that should correct the more speculative approach that, in their view, dominated much biblical scholarship of the 20th century. The rule-based approaches that emerged in the last decades of the 20th century facilitated indeed such a more transparent data-driven analysis. However, in recent digital text analysis rule-based approaches are being replaced by all kinds of Machine Learning techniques. The new techniques provide new exciting potential for biblical studies. However, the highly complicated mechanisms underlying these techniques seem to be moving away from the transparency and insightfulness that the Bible-and-Computer pioneers advocated. These developments raise new questions about the relationship between the outcomes of deep learning applied to text classification, author recognition, source criticism, or linguistic dating in relation to traditional scholarship.

Alessia Pezzella (University of Innsbruck)
"Latin Technical Legal Terminology in Greek: Examples from Egyptian Papyrus Documents"

Latin technical legal terminology is common in Greek juridical texts, especially after the Justinianic reforms. Some earlier examples are found in Greek documentary papyri from late antique Egypt. In the evidence, this terminology is written in Latin script, in Greek script, or in forms that mix both; it features Latin or Greek morphology or a hybrid inflection. The question hence arises of how this terminology can be described from a sociolinguistic perspective: is it borrowing, code-switching, or something in between the two? What is the relationship between languages and scripts? Whereas literary occurrences of the phenomenon have recently been investigated along those lines, occurrences in Greek papyrus documents are mostly unexploited in this respect. The latter evidence, although anecdotal, is important since it offers a complementary perspective on the usage of Latin in juridical texts known via different textual transmission and contributes to our understanding of how Latin could be used in in late antique and Byzantine Egypt. Therefore, the present paper discusses some Latin technical legal words and expressions in Greek papyrus documents and addresses the following questions: What are the language contact phenomena that this terminology implies? Why is it used? Is there a connection with the literary evidence? What can the results tell us about the use of Latin in late antique and Byzantine Egypt?

Fiona Phillips (Corpus Christi, Oxford)
"A Quantitative Approach to Carian-Greek Language Contact"

The dynamics of Carian and Greek in the mid-first millennium BCE remain underexplored. Carian, an Anatolian (IE) language, existed in close contact with Greek from their first attestations in the eighth century BCE into the early Hellenistic period. I hope that a quantitative approach can shed light on the linguistic situation of southwest Anatolia and studies of language contact via epigraphy more generally.

This paper presents the results of a survey of Caria's inscriptions *ca*.800–300 BCE, collated from disparate print and digital sources. It encompasses southern Ionia alongside 'Caria Proper'; considering the former predominantly-Greek-speaking region allows for a comparative analysis with the latter, in turn revealing what is distinctively 'Carian' about Carian epigraphy. It shows that what appears to have been the widespread adoption of the Greek language in central and western Caria was correlated with the trappings of Greek *polis* through the medium of public lapidary epigraphy. The epigraphic production of 'Caria Proper' only outstripped that of Southern Ionia in the fourth century. Epigraphic representation of the Carian language simultaneously declined. As writing on stone was adopted, so too was the Greek language. Nonetheless, numerous exceptions show this phenomenon to have been a trend rather than a rule. Such exceptions should also warn us against assuming epigraphic culture to directly represent spoken language use.

Matthew Robinson (Balliol, Oxford)
"Programming at the Edge of Poetry: a computerised approach to Latin acrostics"

In recent years there has been an explosion of interest in the study of acrostics (words formed by the first letters of a series of lines) in Latin poetry: the more scholars look at the edges of poetry, the more examples of acrostics they find. A key question in recent scholarship concerns how we decide whether or not any of these examples are meaningful, a question usually framed as a distinction between acrostics that are intentional, and those that have occurred by chance; and arguments for intentionality often appeal to statistics and probability. However, for the most part scholars have been working with very small data sets, with criteria for intentionality which lack robust foundation.

However, a programmatic approach to acrostics allows us to map out an entire landscape that was previously only seen in glimpses. For my research I have developed a set of programmes that can uncover acrostics and telestics across the entire corpus of classical Latin poetry, and this enables a range of statistical analysis that has not previously been possible. Now one can not only compare the distribution of acrostics between various authors, but also between their distribution in any particular author and their distribution in a randomly-generated text with the same first-letter frequencies – and the results call into question some commonly-held assumptions about acrostics. In this paper I outline how these programmes work and look at some of the results they have generated.

Eva María Rodrigo Gómez (Hebrew University of Jerusalem)
"Greeks Findings in Syriac Texts: an example in Emmanuel bar Shahhare"

Emmanuel bar Shahhare is a 10th century monk in the Upper Monastery near Mosul, who belonged to the Church of the East. Little is known about his life other than that and his work is yet to be edited, translated, and thoroughly studied. Twenty-eight metrical homilies (memre in Syriac) are attributed to him.

He lived in a fascinating place and moment, at the peak of the Graeco-Arabic translation movement, that produced many renditions of scientific and philosophical Greek texts into Arabic, often through an intermediary Syriac version and almost always done by Syriac translators, some of the most famous ones appertaining to Emmanuel's Church, such as Ḥunayn ibn Isḥāq.

In this communication I intend to explore the Greek influence in Emmanuel's production. The homilies I will study speak about the *Hexaemeron* (the six days of Creation), an ideal topic to introduce "pagan science" to explain the Word of God.

The transliteration of the homilies has been done with eScriptorium, an open software for Handwritten Text Recognition, while the study of the Greek loanwords is aided by a Python code to web scrape all the Syriac words with Greek etymology present in Sedra and then compare them with Emmanuel's texts.

Maroula Salemenou (Classics, Oxford)

"Negotiating linguistic norms in Sappho and Alcaeus: some examples from the Graeco-Roman period in papyri"

Papyri of/on Sappho and Alcaeus from the Graeco-Roman period enable us to source information concerning the correct form of the athematic inflection in -ημι through comparison with the thematic inflection in -εω that is judged similar in the Lesbian verb system due to the presence of shared formal features. Such a comparison had also to be based on literary usage in order to collect the cases in the primary and secondary tradition on Sappho and Alcaeus that form the basis of the *kanónes* for inflectional endings in the Lesbian verb system, and to rank them in terms of statistical probability from the lowest to the highest. The principal criteria of etymology and philological, as well as grammatical, tradition have also been applied to consider which verbal form has to be considered as the norm. Verbal forms and more special verb constructions have been accepted as correct or incorrect Greek, and new restorations have been proposed in the Lesbian corpus, all of which will be presented in a preliminary report.

Jonas Schollmeyer (University of Leipzig)

"Towards a History of Hiatus in Greek Prose"

The Latin term *hiatus* refers to the juxtaposition of a vowel at the end of a word, scanned according to its natural quantity, with the vowel at the beginning of the following word. The examination of whether Greek prose texts allow hiatus is vital for revealing forgeries or interpolations, assessing conjectures and readings, and providing insights into the stylistic shaping of a text or individual passages. However, a comprehensive history of hiatus in Greek prose remains a desideratum, with crucial questions unanswered, such as which hiatus were statistically avoided and to what degree, who avoided them, when, how, why, in which contexts, and by which means.

Most research on this topic dates back to the 19th century and consists mainly of listing instances of hiatus and eliminating them through textual interventions. Additionally, vast text corpora, including many technical writings and large parts of Byzantine prose, have never been studied for their use of hiatus.

With the emergence of digital humanities, there is a great opportunity to revisit this topic and explore its potential fully. In my presentation, I will outline the steps and methods towards establishing a history of hiatus in Greek prose, including initial findings using examples from Gorgias to Polybius, and the obstacles that arise.

The outcomes of this research have implications not only for the evolution of Greek prose style but also for various fields in Classics, such as ancient history, theology, and philosophy, due to the widespread phenomenon of avoiding hiatus.

Matthew I. Swindall (Middle Tennessee State University)
"A.I.-Assisted Papyrology: Integrating Deep Learning into the Scholarly Workflow"

State-of-the-art Optical Character Recognition (OCR) and Handwritten Text Recognition (HTR) approaches are often insufficient to aid the papyrological community due to the mismatch in quality between images in the training datasets versus those of existing texts. Datasets typically consist of images of carefully curated manuscripts which are in pristine condition with clearly legible characters. However, most papyri from the ancient world exhibit numerous forms of damage (holes, tears, stains). Additionally, creating image datasets can be costly, requiring professionally-annotated texts or various paid services. The Ancient Lives Project produced millions of crowdsourced annotations of papyrus fragments. From these annotations, a largescale Greek Character dataset, AL-PUB, was created. Using this dataset, we have developed a pipeline of machine learning tools to assist papyrologists in transcribing ancient Greek papyri. First, a character detection model was created which locates characters in damaged papyri with precision nearing 90%. Second, a character classification model was created which achieved an accuracy of roughly 97%. Third, a line segmentation model using DBSCAN was developed in order to reconstruct the ordering of characters in the manuscript. The result of this pipeline is an automated diplomatic transcription of papyrus fragments. In this presentation, we will discuss and demonstrate this pipeline, then outline our vision for incorporating it into an intelligent, agent-based software that works with the user to assist in the fundamental tasks of critical editing as well as in teaching papyrology.

Marja Vierros (University of Helsinki) (KEYNOTE)
"Digging linguistic data from Greek documentary papyri – Experiences and experiments"

This paper presents a run-through of the project Digital Grammar of Greek Documentary Papyri (ERC Starting Grant agreement No 758481), including information on the linguistically annotated data and digital tools developed for producing and querying the data. I will discuss how they can contribute to our knowledge of the developments of language used in Greek documentary papyri in the multilingual setting of Egypt in the time period from ca. 3rd century BCE to 6th century CE. I will present some preliminary research results with the help of some case studies, and will shortly discuss recent experiments with e.g. word vector models.

David Wilson (SOAS)

“The enriching role of bilingualism in Mesopotamian incantations”

This paper will examine bilingualism in 1st Millennium BCE Mesopotamian therapeutic incantations, focusing on the Udug-ḫul ‘Evil demons’ and Sag-gig ‘Headache’ series of texts to elucidate issues of ancient Sumerian > Akkadian > modern language translation.

Many Mesopotamian incantations are bilingual: they consist of a line of Sumerian followed by a line of Akkadian which is often a clear translation of the Sumerian. There are, however, occasions of nuanced variation in meaning between the two lines. I demonstrate how the interplay between the two languages contributes to the figurative fabric of these therapeutic incantations and creates enriched expressions of pain and suffering in the ANE poetic landscape of illness. Some examples of this are in the form of paronomasia and others exploit ambivalence to metaphorical effect.

Variations between Sumerian and Akkadian, and between different manuscript versions of texts, have become more prominent as ANE researchers have, since the 1980s, published texts in score/partitur form rather than as edited texts with critical apparatus. While the increasing availability of digital transliterations of tablets supports textual variant analysis, the lack of authoritative lexical tools for Sumerian remains challenging.

Greater integration between digital resources can address some of these issues. Nevertheless, the question remains of how best to edit and present these poetic texts in a way that is accurate while also capturing the layers of meaning generated by bilingualism. My paper therefore offers a fruitful comparison with other disciplinary fields that are navigating these same challenges.