DIGITAL CLASSICS ONLINE

More than Names? Challenges and Opportunities for Ancient Named Entity Recognition

Chiara Palladino

Abstract: This paper focuses on the conceptual challenges of Named Entity Recognition and Classification for Ancient Greek and Latin texts. It examines the shifting definitions of 'name' and 'named entity', their changes over time, the overlaps and differences between them, and shows how their use is often flawed by implicit assumptions on naming mechanisms in language and culture. It then offers examples of ancient place-naming practices that may challenge these assumptions, highlighting the limitations of current vocabularies and standards, and pointing to the need for a domain-specific approach to the problem of Named Entities in ancient languages.

Named Entities and Digital Classics

The ability to recognise, identify and analyse information from proper names is a fundamental aspect of humanistic research. In Classical Philology, names of persons and places convey essential information: personal names can lead to an enhanced understanding of social dynamics, character usage in narratives, or family networks. Place names can be extracted and plotted on maps to provide context and to visualise the material spaces a document describes, or they can help better conceptualise travel and movement. Citations of works and authors can provide insight into textual transmission and the circulation of ideas. Referencing obscure names to entries in dictionaries and encyclopaedias can provide an essential aid to reading, teaching, and learning about an ancient document. Furthermore, the indexing and association of textual information with external metadata, such as findspots, editors and collections can support more systematic research into material artefacts such as papyri, coins and inscriptions.

The automatic extraction and classification of names is conducted through Named Entity Recognition (NER), a basic task of Natural Language Processing (NLP). At the most basic level, NER consists of the parsing of a text to extract relevant strings, which represent names. The importance of NER has long been recognised in Computer Science, as Named Entities are more or less static strings in language and can be used as anchors to improve on various text processing tasks, such as morphosyntactic and part-of-speech tagging. The recognition and indexing of names, however, is also a fundamental part of Classical Philology. Organising, cataloguing, and indexing book titles, author names, and place information constitute one of the chief occupations of the philologist since antiquity.

In recent years, the revolution of Transformers and Large Language Models has had a significant impact on the computational study of the ancient world: the emphasis has been very much on linguistic analysis, handwriting recognition, and translation, but there is increasing attention towards extraction

and classification tasks.[1] Various NER models are currently available for Latin, Ancient Greek, and other premodern languages, with significantly improved performances on existing benchmarks.[2]

However, these methods face structural challenges that are typical of the computational processing of ancient documents. Some of the problems depend on the externalities of available technologies, which may include outdated edition formatting, noisy OCR output, unstable orthographies or spellings, or lack of unified Unicode standards.[3] More specifically, performance issues in NER models have been associated with the inability to extract outliers like names in foreign languages, or with inconsistencies in the detection of string boundaries in the case of multi-word names. Significant performance drops happen regularly when such models are tested on out-of-domain texts, which indicates overfitting in training and fine-tuning: this problem is related to the paucity of available high-quality annotated datasets and reliable guidelines.[4]

Finally, the lack of domain-specific tagsets is a challenge to research in this area: the labels used to classify entities according to semantic typologies are predominantly adapted from modern NLP tagsets, which are often used in domains that have very little in common with ancient documents. Considerable progress has been made in the domain of citation of ancient authors and works,[5] which can be annotated and extracted through carefully designed guidelines. Palladino et al. proposed an experimental tagset designed specifically for the annotation in Latin and Greek, with the intent of ensuring a basic level of interoperability across annotated corpora.[6] However, there is no agreed-upon strategy for handling named entities as structured information in ancient texts: in a sector where the lack of annotated datasets is an increasingly important problem, reconciliation strategies are often employed to ensure interoperability and data exchange across projects and to support fair evaluation of NER outputs.[7]

This situation reveals a fundamental lack of consensus on the definition of *ancient named entities*, and what criteria can be used to recognise them. In fact, what exactly constitutes a Named Entity, and to what extent it overlaps with the grammatical notion of name, is far from established. In this paper, we will more closely examine these competing notions and their ties to linguistic and cultural definitions. We will mainly consider toponyms in Ancient Greek and Roman sources, as a case study where cultural mechanisms of naming and transmission introduce a layer of complexity that helps question our preconceived notions of what constitutes a name, as well as problematising the assumptions underpinning extraction and classification tasks.

What is a Name?

The definition of what constitutes a name is obviously complex, and it has been studied in an impressive number of fields, from philosophy to logic, to linguistics, to anthropology. In fact, the defining criteria of what constitutes a name in language are very nuanced and full of potential contradictions.

We usually refer to names in language as something distinct and different from common nouns, and with a certain degree of specificity. In grammar, some criteria are commonly considered as defining proper names: capitalisation, lack of lexical meaning/semantic emptiness, morphosyntactic regularity

1 Sommerschield et al. (2023).

2 Beersmans et al. (2023).

3 Ehrmann et al. (2024).

4 Palladino / Yousef (forthcoming).

5 Romanello / Najem-Meyer (2022); Berti (2023).

6 Palladino et al. (2024).

7 Palladino / Yousef (2024).

(lack of inflection and determiners), untranslatability, and absence in traditional dictionaries. In general, these criteria point to the seemingly unique status of names as static strings of language that do not have a stable lexicographical definition like common nouns.

This definition is usually enriched by an idea that derives from the philosophy of language, the definition of name as a *rigid designator* that points to a *referent*. The concept owes considerable debt to Kripke:[8] differently from a common noun, a name is considered to point unambiguously to a uniquely identifiable object, the referent. For instance, while the noun 'city' indicates a class of objects, 'Athens' unambiguously denotes an individual instance of that class, that is, a specific thing in the world.

The concept of Named Entity is obviously much more recent. Officially, the idea of 'Entity Expression' emerged during the MUC-6 Evaluation Campaign in 1995, even though already in 1991 Lisa F. Rau designed a method to extract company names from texts, a task commonly understood as part of NER today.[9] The chief idea was to extract text strings that could be identified as entities of interest, and further disambiguate them using labels, or tags, that identified their type. In 1995, entities of interest were grouped as 'Entity Name Expressions', identified as persons, organisations and locations, 'Numeric Expressions' and 'Time Expressions'. Events, relationships, and coreference resolution were progressively added to the task, although today they are often considered more advanced applications. The benefits of extracting and classifying Named Entities were numerous, from a better understanding of context to improvements in other areas of language processing, like morphosyntactic parsing or translation.

However, what exactly constitutes a Named Entity and how to recognise it are questions that have shifted significantly over the years, and changes have occurred in almost every major evaluation campaign since 1995.[10] Named Entities never received an authoritative linguistic definition, contributing to the mistaken assumption that they may overlap with proper names: since the very inception, however, it is evident that the idea of Named Entity was broader and more flexible, but also vaguer.

The most stable defining criterion for a Named Entity seems to be the idea of rigid designator in the Kripkean sense.[11] This is also the assumption behind Entity Linking, which is based on the idea that a Named Entity extracted from a text can be associated with the corresponding referent, as it appears recorded in an ontology or authority list, through the operation of *semantic annotation*. However, the continuous redefinition of the domains of interest for NER, alongside a considerable expansion in label vocabularies, led to a certain differentiation in what the designator represents, increasingly problematising the Kripkean notion.

The relationship between *name*, *entity*, and *referent* is not a stable one. In its original formulation, anything could potentially be considered a referent: 'gold' or 'whale', for example, are specific instances of more general classes of objects, but they still designate something with specific properties and characteristics without being proper names in the grammatical sense.[12]

The question, therefore, becomes one of granularity. In their seminal work on NER, Nouvel et al. show that, while the idea of denoting something unique is somehow common to all definitions, the mechan-

8 Kripke (1996).

9 Marrero et al. (2013).

10 See Appendix 5 in Nouvel et al. (2016) for a non-exhaustive list.

11 Note, however, that Nadeau / Sekine (2007) referred to potentially *one or many* rigid designators.

12 Marrero et al. (2013). Additional problems relate to the notion of coreference, where pronouns, synonyms, and other discourse elements may also be associated with unique referents. For the sake of space, we will follow recent trends in NER that consider coreference as a separate task, and we will limit the present discussion to names.

isms through which this is achieved are heterogeneous.[13] This is one of the most important differences between the grammatical definition of name and the concept of Named Entity. Descriptive expressions are often considered NEs, provided that they point to a referent: for example, the string 'the daughter of Augustus' contains one proper name (Augustus), but has another identifiable and stable referent, Julia the Elder. Therefore, the *entity* it refers to is different from the *name* it contains (or there are two entities, but only one name). In other cases, a name in the grammatical sense may not be an entity at all: in the sentence 'Athens is a place-name', 'Athens' is not an identifiable city but *the name of the name* Athens, opposed to 'the name Alexandria' or 'the name Rome'. Thus, whether there is a referent or not depends very much on the definition of what constitutes a 'real object'.

Other categories are inconsistently grouped under the definition of name, Named Entity, or neither: patronymics, ancestry names, nicknames, political roles tend to have limited importance in traditional NLP, but are extremely important in literary studies and are sufficiently stable in their specific context that they may function as rigid designators. In situations of coreference, different names may point to the same referent, but be perceived as having different meanings or even point to culturally different objects: famously, this is the case of Frege's puzzle, where the 'morning star' and the 'evening star' have the same referent (the Planet Venus), but differ in their meaning to the point of being perceived as different versions of Venus (the planet as it appears in different locations in the sky a few months apart) and the phrase 'the morning star is the evening star' is not a tautology but an expression contributing new knowledge.

Fort et al. define Named Entities based on the concepts of *referential unicity* (contextual unicity of the referent), *referential autonomy* (should be sufficient to identify the referent), *denominational stability* (more regular and less numerous than other noun phrases), and *referential relativity* (the referent is considered relatively to a domain model).[14] The notion of referential autonomy, as discussed above, does not necessarily mean that a Named Entity is a name, but rather that the text string is sufficient, without additional context, to identify a specific object. On the other hand, referential unicity and relativity imply a consideration of *domain specificity*. The extraction and classification of Named Entities must be useful for a goal, rather than being a text processing exercise: therefore, the application of the task should be preceded by the delimitation of the elements of interest for the domain under investigation. The idea of Named Entity becomes, therefore, increasingly domain-dependent: as different knowledge domains may have different ideas of what constitutes a string of interest, it becomes a more flexible concept than the grammatical notion of proper name.

Following this trend, tags and vocabularies to classify Named Entities have grown exponentially to accommodate increasingly diverse domains, leading to tagsets of more than 200 labels and to strategies of annotation that divide Named Entities into many different components, aiming to achieve very fine semantic granularity.[15] Another approach to the problem has been suggested using nested entities, that is, the construction of named entities made up of multiple names in a hierarchical 'nested' structure. Strategies of this complexity become very close to ontologies, as they require multiple levels of internal ramifications.

On the other hand, the advent of Large Language Models and Generative AI has highlighted the lack of consistency in the development of datasets. Machine Learning, and particularly language models, require well-defined benchmarks to achieve reliable results in training and evaluation. Therefore, recent research has emphasised the need for harmonisation across labelling systems, in order to decrease

13 Nouvel et al. (2016).

14 Fort et al. (2009).

15 Sekine et al. (2002).

redundancy and ambiguity in datasets.[16] Thus, NLP is divided between these two demands, and this becomes more impactful in cases where the quantity and availability of data is limited.

There seems to be a tendency to trust the annotator's common sense to know what constitutes a Named Entity in each context: ultimately, the individual's judgement is considered the best criterion. This, however, is a problematic assumption, based on the idea that the concept of name and naming mechanisms are somehow intuitive and almost cultural universals. Not only this is demonstrably untrue, but it is especially dangerous for contexts that are situated far away in time and space from the annotator. The need for consistent datasets requires a careful definition of what constitutes a Named Entity, and the instability of this concept does not provide easy solutions.

In the section that follows, we will focus on ancient place-names, or toponyms, to illustrate how naming practices pose crucial challenges to the definitions described above. We focus on place-names specifically, not only to keep within the space allotted by this paper, but also because place-naming is a very good example of a practice with cultural specificity, as the context of creation of a name affects its manifestations in documents. There is growing scholarship on personal naming practices in the ancient world, but the study of toponymic practices and cultural transmission of place-names is often limited to etymology or individual case studies.[17] Place-naming as a cultural practice, while abundantly investigated in other fields, is still a novel area of investigation for Classical Antiquity.

Ancient Names, Modern Named Entities

Names and Place-Naming Practices

The operation of naming something obviously bears a degree of intentionality, even though the original meaning of a name may be lost in time. Place-naming is a social act common to all human cultures. It is impossible to detach place-naming from the human experience of place-making, that is, the process of understanding, conceptualising, and communicating spatial knowledge. Place-names stabilise features in the environment around which human activity is organised, but also function as linguistic tools to store and transmit geographical information.[18] Far from being a simple top-down imposition of a conventional model of reality, place-names are intimately connected with the material world they define, but also with local knowledge and memory.[19]

The standard grammatical definition of proper name is problematic for ancient, indigenous, and non-Western languages, and sometimes it directly contradicts the dynamics of place-naming. Criteria such as lack of inflection or determinants, as well as untranslatability or exclusion from dictionaries, are generally unsatisfactory. Likewise, dictionaries have an inconsistent approach to the inclusion of proper names, particularly place-name derivatives such as ethnonyms and demonyms.

It hardly needs mention that capitalisation is an unreliable mechanism to distinguish proper names in ancient texts. Capitalisation did not consolidate in Greek and Latin scripts as an orthographic convention to mark proper names until the Middle Ages, and even then, it was used inconsistently. This means that, in the manuscript tradition, capitalisation is not the prevalent convention to mark names, which are often not distinguished at all from the rest of the text. However, since we work mainly with modern critical editions, the editorial intermediation has already set out a text where certain distinc-

16 Palladino / Yousef (2024).

17 On personal names, see the recent contributions of De La Escosura Balbás et al. (2024); Bonnet (2024). On toponymy, notable etymological studies are Georgacas (1959); McDonald (1958).

18 Eades (2017).

19 Palladino (2023).

tions were made by typographic conventions or conscious choices that are almost never documented in the critical apparatus.

This is most visible in cases where geographic features originate from, or are included in, a toponym. One of the most basic place-naming mechanisms is using a notable feature to name an area, blurring the boundary between naming and descriptive function. While the assumption is that the name is going to progressively lose its lexical meaning and become associated with the referent, this is not always the case. Furthermore, what is valid for modern perceptions is not necessarily similar in ancient contexts. This leads to some notable inconsistencies in the editorial tradition of proper names.

One example of this is the Isthmus of Corinth. This toponym is often indicated in texts through the common noun 'ἰσθμός', which appears capitalised in modern editions to distinguish it from other straits located elsewhere (with notable inconsistencies, where even within the same edition the word is occasionally present in lowercase when referred to the Corinthian isthmus: see for example Paus. 2,1,3; Diod. 12,59,1).[20] This, however, is purely a modern convention based on modern associations with a historically important location: the correct identification of the isthmus does not rely on the word itself, but on the notion that the geographical context of the passage is known to the reader.

The case of the Pillars of Heracles ('στῆλαι Ἡράκλειαι') is the opposite: the common noun is never capitalised in modern editions even though it indicates the toponym (which is, however, regularly capitalised in translations). Two important exceptions are Philostr. Ap. 4,47 and one passage of the *Geography* of Strabo (15,1,6), where we find 'Στῆλαι'.[21]

The Acropolis of Athens is generally considered a place-name today: however, the term 'ἀκρόπολις' simply denotes a spatially identifiable part of a polis, equivalent to the common noun 'citadel'. In fact, the acropolis of Thebes has its own name, Cadmea. A similar case is 'τὰ μακρὰ τείχη', which appears always lowercase (e.g. Xen. Hell. 2,3,11; 4,4,18; Thuk. 1,107; 5,26; etc.),[22] but is consistently capitalised in translations because it indicates a precise object in the mind of a modern reader, the Long Walls of Athens.

These cases are problematic, because they only function as rigid designators within a shared spatial framework (such as the city of Athens), and the uniqueness of their referent is purely based on a geographical characteristic. They most definitely were *places*, but their usage does not fit the strict linguistic definition of *name*, as they are used with their original lexical meaning.

The idea of lack of lexical meaning implies that a place-name does not carry information about the object being designated: the 'White House' does not designate a house that is white, while a noun generally gives information about the type of object being described ('house'). In the case of toponyms, this goes hand in hand with the phenomenon of grammaticalisation, where a toponym with a specific etymology and clear origin progressively loses that association, becoming devoid of meaning.

The fact that toponyms grammaticalise, however, is not sufficient to consider them meaningless. This is especially problematic for indigenous toponymic systems, where place-naming practices harness the whole descriptive power of language and toponyms are designed to provide information about the places they refer to.[23] Fields like critical toponymy recognise place-names and place-naming as cultural practice to consider within both connotative and denotative functions, being at the same time util-

20 See for example: Spiro (1903); Jones / Wycherley (1955); Musti (1986); Fischer / Vogel (1888–1906); Oldfather (1933–1967).

21 Meineke (1877); Kayser (1870).

22 We cite here the authoritative Oxford editions, but various other available editions have been consulted: Marchant (1900); Jones / Powell (1942).

23 Keith Basso demonstrates this powerful mechanism in Western Apache place-names, which include examples like 'Green Rocks Side By Side Jut Down Into Water', or 'Gray Willows Curve Around A Bend'; Basso (1996), 23.

itarian and symbolic.[24] In the ancient world, the strong relationship between the geography of a place and the meanings associated to it by its inhabitants are highlighted by Strabo in his proem to the *Geography* (2,5,17): in addition to geographic features, the traditions and values associated to them by people are imposed, but end up becoming almost natural constituents of a place, although (differently from physical characteristics) they are subject to change through time.

Ancient toponyms may be lexically meaningless in modern languages, but they may not have been in their original usage. The expression 'συμπληγάδες (πέτραι)' (Apollod. 1,9,22) illustrates the origin of a place-name through this explanatory/descriptive mechanism. In the passage, Phineus describes the dangers of the route to the Argonauts, including the 'clashing rocks', an area continuously tormented by the wind and the sea. The expression is used to describe a feature of the landscape in a meaningful way. 'Symplegades', however, is also the actual toponym of the area, undoubtedly attested as such in ancient literature (and even capitalised in the English translation of this passage).[25] This ambiguity reflects the complicated mechanisms of place-naming, where the boundary between toponym and meaning is not always clear-cut.

Whether through actual etymology or through transmission practices, ancient toponyms bear the traces of this process of landscape conceptualisation: for example, the names of the minor seas of the Mediterranean were linked to the surrounding areas or islands ('Tyrrhenian', 'Phoenician'), environmental characteristics (e.g. 'Pontos Euxeinos', 'hospitable', vs. 'Pontos Axeinos', 'inhospitable'), local mythologies, or even navigational information, such as the dangerous character of an area ('Skyllaion') or the local winds to which it was exposed ('Zephyrios Limen').[26] Although many toponyms progressively lose association with their original etymologies, this is often counterbalanced by the consolidation of folk etymologies reflecting dynamics of degrammaticalisation: far from demonstrating lack of meaning, in the ancient world the attestation of different traditions on the origins of toponyms show the strategic importance attached to places. So, the many competing origin stories that associated the Aegean Sea with various local heroes and traditions reflect a desire by various communities (notably, but not exclusively Athens) to claim a strategic natural affinity with that space.[27]

The use and understanding of toponyms require a shared cultural background, but also some knowledge of local spatial configurations. Thus, what Ancient Greeks and Romans considered as place-names may have different answers. In some ways, this is a translation problem: there is no perfect cultural equivalent for a word or expression across different languages, and this extends to naming practices. What defines a name is culturally specific, rather than a universal: therefore, different cognitive and linguistic mechanisms may point to different definitions of what a name and its components are.

Place-Names as Named Entities

This descriptive and additive nature makes toponyms an interesting challenge for Named Entity Recognition. As we have seen above, Named Entities are defined under broader terms than names, and they rely on more flexible and domain-specific criteria to define strings of interest. The individuation of a string of interest relies on the ability to recognise it as a designator for a unique object in the world: this idea, however, relies on some level of shared knowledge.

One of the key assumptions behind Named Entities is that there is knowledge that the referent exists, but this is not always the case for the ancient world, where much contextual information is vague or lost to us, and we cannot always be sure whether we can associate a word in a document with a spe-

24 Choo (2023).

25 See for instance the famous Frazer translation (1921), and more recent translations in English by Hard (1997) and Smith / Trzaskoma (2007). See also the most recent authoritative Italian translation by Sarpi / Ciani (1996).

26 For these examples, see especially Morton (2001), 70ff.

27 Ceccarelli (2012).

cific individual or place. This does not mean that they did not originally exist or were not considered real.

Furthermore, place-names are, by definition, distributed and indefinite:[28] while it is quite easy to identify the physical boundaries of the individual named 'Augustus', it is extremely difficult to precisely define where 'Roma' begins or ends, and those boundaries change through time and context. Defining what makes anything a 'place' is a complicated matter. For example, defining a place on the basis of geographical location or cartographic coordinates is deeply problematic for the ancient world. Places may also change their physical location or configurations, sometimes significantly, in the course of a very long chronological span.[29] Thus, the concrete things denoted by a place-name could be identified in very different ways, depending on context: the referent of a place-name and its characteristics are not necessarily stable.

As we have seen above, Named Entities tend to include descriptive naming mechanisms ('the daughter of Augustus') under their definition. However, because place-naming is by definition descriptive and relational, it is much harder to understand when a string of text represents a descriptive place-name and when it is a simple noun phrase, especially when there is no contextual information or any evidence of consolidated usage patterns. Because different languages have very different place-naming practices, virtually any sort of geographical description could be considered a place-name, as long as it designates a specific and identifiable feature in a unique way and its usage can be demonstrated across a variety of contexts.[30] In other words, in many cases it is very difficult to understand whether a *spatial entity* is also a *named entity*.

This represents one of the biggest challenges for the automated extraction of place-names, the detection of entity boundaries: in many cases, it is very difficult to establish where a place-name begins and ends. The most common example is instances where a proper name in the conventional sense appears with descriptive attributes or common nouns, denoting a geographical feature, such as 'Νεῖλος ποταμός' or 'Κάσιος ὄρος', or have the function of further specifying a place in implicit contraposition with a different one, as in the case of 'κάτω Αἴγυπτος', '*mare Superum*', or '*Germania Inferior*'. These cases also challenge conventional morphosyntactic annotation, because different strategies are adopted to mark proper names and their constituents.[31]

More complex expressions like 'Ἡράκλειαι στῆλαι αἱ ἐν τῇ Εὐρώπῃ' (Ps. Skyl. 1) or 'Ἰβηρικὴ παραλία' (Strab. 3,4,16) certainly denote *spatial entities*, because they can be associated with a specific geographic area that is precisely recognizable and cannot be confused with any other. However, the problem is to understand whether they are perceived as *place-names*. In these cases, the attestation of regular usage in the tradition is the only discriminating criterion that can be adopted. Even so, some instances may put into question our own lack of evidence and our familiarity with the context. Ancient ethnographic practices, for example, often use relational expressions to denote specific subgroups: the Ethiopians are a particularly interesting example, where the regularity of usage to indicate distinct geographic areas based on relative location ('ὑπὲρ Συήνης', 'πρὸς/ὑπὲρ Αἰγύπτῳ', 'ὑπὲρ Μαύρους', etc.) suggests that these were fixed expressions, indicating different subgroups often marked in modern commentaries as 'Eastern' and 'Western' Ethiopians.[32]

---

28 Eades (2017).

29 Georgacas (1959).

30 Clearly, personal names are not totally devoid of the issue. Strings like 'the winner of the 60th Olympiad' or 'the wife of Augustus' nephew' have their own share of problems. In the case of place-names, however, the question has a cognitive implication: these expressions may not have been considered as noun phrases, but as actual *place-names*, in the same way as 'Rome' or 'Athens'.

31 Because of the strong relation between toponym and landscape terms, this is also a translation problem, as languages differ wildly in the ways in which they define landscape features such as 'mountain' or 'sea'.

In computational processing, boundaries need to be clearly defined and cannot be vague. The definition of boundaries may also affect subsequent tasks, such as the classification via semantic vocabularies. The inclusion or exclusion of descriptive words within the name may mean a change in label: cases such as 'city of the Gaditanes' ('Γαδιτανῶν πόλις'), 'pillars of Heracles' ('Ἡράκλειαι στῆλαι'), 'temple of Juno' ('*Junonis templum*'), 'altar of Apollo' ('βωμὸς Ἀπόλλωνος') are clear instances of this, where the *name* appearing in the text actually differs from the type of the *entity* being talked about.

The task of semantic labelling is peculiar to the treatment of Named Entities in the computational space, as it supports further disambiguation and processing, and it helps define extraction patterns. However, the classification of an entity is never completely devoid of the surrounding context.

Geographical features like constellations, winds, and compass points tend to be used differently in ancient and modern traditions. Wind names, for example, may be used in the ancient world to indicate actual winds or to express cardinal directions metonymically ('towards Zephyrus'), but winds are also regularly personified, calling into question the entire usage of capitalisation in modern editions.[33] Similarly, ancient Mediterranean cultures have a tendency to personify (or, more precisely, deify) rivers, as representations of important seasonal phenomena (the Nile and the Egyptian god Hapi) or mythical entities (the Styx in the Greek tradition). Spatial features or natural forces are often associated with superhuman beings of various kinds, but at least in part retain their material characteristics, to the point where it becomes extremely tricky to classify them using modern definitions: this difficulty is an indication of the challenges behind the idea of a uniquely identifiable referent, when in practice, the referent for the name 'Styx' is a dualistic object.

A linguistically interesting example is provided by phenomena of metonymy, that is, the practice of using the name of a place to refer to its inhabitants ('France won the World Cup') or vice-versa: the former is considered typical of modern languages, while the latter is much more common in Ancient Greek toponymic practices.[34] So much so that the 'people-for-place' mechanism is evident in the etymology of certain Greek toponyms, which derive from the ethnonym of their inhabitants. The most famous case is Delphi, where the toponym 'Δελφοί' is actually the name of the local population. However, 'Δελφοί' in its original ethnonymic sense did not fall out of use, and it is attested in texts to indicate either the Delphian people or the place of Delphi.[35] While this distinction can sometimes be clarified from context and some occurrences are used unambiguously for the people or the place, the decision is often left to the scholar's interpretation. This phenomenon, however, is a core practice of ancient place-naming, and it may well be that even an ancient author would not be able to tell the difference – and it may not have mattered.

Conclusions

A substantial part of the operation of recognising and isolating Named Entities relies on the existence of an implicit common interpretive framework, where there is fundamental agreement on what constitutes a distinct piece of information. In the case of ancient documents, however, the uncertainty of the surrounding context makes this assumption fragile. The nuances involved in the formation of ancient toponyms, ethnonyms, epithets, and other named features like winds and personified natural forces, challenge existing definitions at many levels: historical, interpretive, linguistic, and operational. Thus,

32 See Agathem. 2,7; Hdt. 7,70,1; Paus. 1,33,4; 1,33,5; 6,26,2; Strab. 1,2,28; 2,3,8; 17,1,53.

33 Shipley (2021).

34 Poibeau (2006).

35 Kron et al. (2019).

ancient place-names force us to contextualise our definitions as the product of a cultural and cognitive framework, and help us problematise ideas of 'name' and 'named entity'.

Clearly, one does not need to reconstruct a toponym's etymology or history to recognise it as a toponym. However, computational processing requires a remarkable degree of precision in the definition of what constitutes information of interest, and clearly defined criteria are paramount to the creation of guidelines and datasets. For this reason, the study of how an ancient community conceived naming practices and the perceptions attached to those names provides useful insights and supports a better definition of the matter under investigation.

On the other hand, when it comes to the concrete study of a text, it is important to emphasise that the simple extraction of names is hardly enough. To support serious scholarly analysis of ancient place information, it is necessary to further link toponyms to location data and gazetteer information through Entity Linking,[36] but also to combine this information with the analysis of morphosyntax, sentiments, and ideas associated with such occurrences.

Amidst all these challenges, the more flexible conceptual nature of Named Entities presents an opportunity, rather than a limitation. The idea of NE does not rely on rigid conceptual boundaries but focuses on domain-specific definitions within the context of application. Therefore, it provides new ways to look at entities as structured information in texts. We have more freedom to consider named entities through the framework of culturally situated practices, such as naming and spatial conceptualisation. Thus, rather than Named Entity Recognition, it would be more productive to think in terms of Information Extraction within specific knowledge domains.[37]

Leveraging on domain knowledge, therefore, is fundamental for the definition of methods of computational analysis. The challenge is to design vocabularies and standards that take the cultural and linguistic specificity of ancient naming practices into account.[38] This scholarly endeavour is, at its core, philological, but goes hand in hand with the opportunities of automated processing and digital treatment of ancient documents, in order to allow for in-depth investigation to a bigger and more ambitious scale.

36 Beersmans et al. (2024).

37 For example, Romanello and Najiem Meyer (2022) leverage on the notion of 'Knowledge Entity' readapted from Zhang et al. (2021) for the extraction of citations of Classical works.

38 Broux (2015).

Sources

Online Sources

Open Greek and Latin, Perseus Digital Library, Scaife Viewer, https://scaife.perseus.org/ (last access 04.09.2025).

ΛΟΓΕΙΟΝ, https://logeion.uchicago.edu/ (last access 04.09.2025).

TLG, Thesaurus Linguae Graecae, https://stephanus.tlg.uci.edu/ (last access 04.09.2025).

Editions

Fischer / Vogel (1888-1906): K. T. Fischer (post I. Bekker & L. Dindorf) and F. Vogel, Diodori bibliotheca historica, 5 vols., 3rd edn., Leipzig 1888–1906 (repr. 1964), retrieved from: http://stephanus.tlg.uci.edu/Iris/Cite?0060:001:1447586 (last access 28.07.2025).

Frazer (1921): James G. Frazer (ed.), Apollodorus, The Library, Volume I: Books 1–3.9, Cambridge (Mass.) 1921.

Hard (1997): Robin Hard (ed.), Apollodorus, The Library of Greek Mythology, Oxford / New York 1997.

Jones / Powell (1942): H. S. Jones / J. E. Powell, Thucydidis Historiae, 2 vols., Oxford 1942 (repr. 1963).

Jones / Wycherley (1955): W. H. S. Jones / R. E. Wycherley, Pausanias' Description of Greece, 5 vols, Cambridge (Mass.) / London 1955.

Kayser (1870): C. L. Kayser, Flavii Philostrati opera, vol. 1, Leipzig 1870 (repr. Hildesheim 1964), retrieved from: http://stephanus.tlg.uci.edu/Iris/Cite?0638:001:312314 (last access 28.07.2025).

Marchant (1900): E. C. Marchant, Xenophontis Opera Omnia, vol. 1, Oxford 1900 (repr. 1968).

Meineke (1877): A. Meineke, Strabonis geographica, 3 vols., Leipzig 1877, retrieved from: http://stephanus.tlg.uci.edu/Iris/Cite?0099:001:1841753 (last access 28.07.2025).

Musti (1986): Pausania, Guida della Grecia. Libro II. La Corinzia e l'Argolide. Testo e traduzione a cura di D. Musti, Milano 1986 (repr. 2008).

Oldfather (1933–1967): C. H. Oldfather, Diodorus Siculus. Library of History, Cambridge (Mass.) / London 1933–1967.

Sarpi / Ciani (1996): Apollodoro. I miti Greci (Biblioteca). A cura di P. Sarpi. Traduzione di M. G. Ciani, Milano 1996 (repr. 2013).

Smith / Trzaskoma (2007): Apollodorus' Library and Hyginus' Fabulae. Two Handbooks of Greek Mythology. Translated, with Introductions, by R. Scott Smith and Stephen M. Trzaskoma, Indianapolis / Cambridge 2007.

Spiro (1903): F. Spiro, Pausaniae Graeciae descriptio, 3 vols., Leipzig: Teubner, 1903 (repr. 1:1967), retrieved from: http://stephanus.tlg.uci.edu/Iris/Cite?0525:001:209605 (last access 28.07.2025).

References

Basso (1996): K. H. Basso, Wisdom Sits in Places: Landscape and Language Among the Western Apache, Albuquerque (NM) 1996.

Beersmans et al. (2023): M. Beersmans / E. de Graaf / T. Van de Cruys / M. Fantoli. Training and Evaluation of Named Entity Recognition Models for Classical Latin, in: Proceedings of the Ancient Language Processing Workshop (ALP 2023), Shoumen (Bulgaria) 2023.

Beersmans et al. (2024): M. Beersmans / A. Keersmaekers / E. De Graaf / T. Van De Cruys / M. Depauw / M. Fantoli, "Gotta catch 'em all!": Retrieving people in Ancient Greek texts combining transformer models and domain knowledge, in: Proceedings of the 1st Workshop on Machine Learning for Ancient Languages (ML4AL 2024), Bangkok / online 2024, 152–164.

Berti (2023): M. Berti, Named Entity Recognition for a Text-Based Catalog of Ancient Greek Authors and Works, online 2023, https://zenodo.org/records/8108058 (last access 04.09.2025).

Bonnet (2024): C. Bonnet / R. Häussler (transl.), The Names of the Gods in Ancient Mediterranean Religions, Cambridge 2024.

Broux (2015): Y. Broux, Graeco-Egyptian Naming Practices: A Network Perspective, Greek, Roman, and Byzantine Studies 55 (2015), 706–720.

Ceccarelli (2012): P. Ceccarelli, Naming the Aegean Sea, Mediterranean Historical Review 27 (2012), 25–49.

Choo (2023): S. Choo, Assessing the Validity of Critical Toponymy Perspectives for Understanding Human Perception of Places: An Analytical Framework, in: G. O'Reilly (ed.), Place Naming, Identities and Geography, Cham 2023, 29–50.

De La Escosura Balbás et al. (2024): C. de la Escosura Balbás / A. Kurilic / G. E. Rallo (eds.), Name and Identity. Selected Studies on Ancient Anthroponymy through the Mediterranean, Oxford 2024.

Eades (2017): G. L. Eades, The Geography of Names: Indigenous to Post-Foundational, London / New York 2017.

Ehrmann et al. (2024): M. Ehrmann / A. Hamdi / E. L. Pontes / M. Romanello / A. Doucet, Named Entity Recognition and Classification in Historical Documents: A Survey, ACM Computing Surveys 56 (2024), 1–47.

Fort et al. (2009): K. Fort / M. Ehrmann / A. Nazarenko, Towards a Methodology for Named Entities Annotation, in: Proceedings of the Third Linguistic Annotation Workshop on – ACL-IJCNLP '09, Suntec (Singapore) 2009, 142–145.

Georgacas (1959): D. J. Georgacas, A Contribution to Study of Greek Toponymy, Names. A Journal of Onomastics 7/2 (1959), 65–83.

Kripke (1996): S. A. Kripke, Naming and necessity, Oxford / Cambridge 1996.

Kron et al. (2019): C. Kron / W. L. Little / J. C. Wolfe / P. Ajaka / B. O. Allen / C. Brown / M.-C. H. de Marneffe / M. Elsner / M. D. Grioni / B. D. Joseph / A. B. Kessler / H. F. Young, What's In a Name? Issues in Named Entity Recognition, Paper to be presented at Annual Meeting of American Name Society, New York 2019, https://bpb-us-w2.wpmucdn.com/u.osu.edu/dist/4/27964/files/2016/01/ANSabstract-FINAL-2kf9qiy.pdf (last access 28.07.2025).

Marrero et al. (2013): M. Marrero / J. Urbano / S. Sánchez-Cuadrado / J. Morato / J. M. Gómez-Berbís, Named Entity Recognition: Fallacies, Challenges and Opportunities, Computer Standards & Interfaces 35 (2013), 482–489.

McDonald (1958): W. A. McDonald, Early Greek Attitudes toward Environment As Indicated in the Place-Names, Names. A Journal of Onomastics 6 (1958), 208–216.

Morton (2001): J. Morton, The Role of the Physical Environment in Ancient Greek Seafaring, Leiden / Boston 2001.

Nadeau / Sekine (2007): D. Nadeau / S. Sekine, A Survey of Named Entity Recognition and Classification, Lingvisticae Investigationes 30 (2007), 3–26.

Nouvel et al. (2016): D. Nouvel / M. Ehrmann / S. Rosset, Named Entities for Computational Linguistics, London / Hoboken 2016.

Palladino (2023): C. Palladino, Not the Same Landscape. Rediscussing Digital Approaches to Spatial Knowledge Systems, in: C. Palladino / G. Bodard (eds.), Can't Touch This: Digital Approaches to Materiality in Cultural Heritage, London 2023.

Palladino et al. (2024): C. Palladino / M. Fantoli / E. de Graaf / M. Berti / M. Romanello / T. Yousef / M. Beersmans / T. Gheldof / L. Soffiantini / E. Litta Modignani Picozzi, Experience and Challenges with Named Entities – Workshop at DHBenelux 2024, Leuven / online 2024, https://zenodo.org/records/11366870 (last access 04.09.2025).

Palladino / Yousef (2024): C. Palladino / T. Yousef, Development of Robust NER Models and Named Entity Tagsets for Ancient Greek, in: R. Sprugnoli / M. Passarotti (eds.), Proceedings of the Third Workshop on Language Technologies for Historical and Ancient Languages (LT4HALA) @ LREC-COLING-2024, Torino 2024, 89–97.

Palladino / Yousef (forthcoming): C. Palladino / T. Yousef, Named Entity Recognition in Classical Languages: Two Approaches, in: E. De Graaf / A. Keersmaekers / S. Stopponi / S. Peels-Matthey (eds.), Computational Approaches to Ancient Greek and Latin, forthcoming.

Poibeau (2006): T. Poibeau, Dealing with Metonymic Readings of Named Entities, in: Proceedings of the 28th Annual Conference of the Cognitive Science Society (CogSci 2006), Vancouver 2006, 1962–1968.

Romanello / Najem-Meyer (2022): M. Romanello / S. Najem-Meyer, Guidelines for the Annotation of Named Entities in the Domain of Classics, online 2022, https://doi.org/10.5281/zenodo.6368101 (last access 04.09.2025).

Sekine et al. (2002): S. Sekine / K. Sudo / C. Nobata, Extended Named Entity Hierarchy, in: M. González Rodríguez / C. P. Suarez Araujo (ed.), Proceedings of the Third International Conference on Language Resources and Evaluation (LREC'02), Las Palmas 2002.

Shipley (2021): D. G. J. Shipley, Sun, Sea, and Sky: On Translating Directions (and Other Terms) in the Greek Geographers, in: E. Boutsikas / S. C. McCluskey / J. Steele (ed.), Advancing Cultural Astronomy: Studies In Honour of Clive Ruggles, Cham 2021, 105–136.

Sommerschield et al. (2023): T. Sommerschield / Y. Assael / J. Pavlopoulos / V. Stefanak / A. Senior / C. Dyer / J. Bodel / J. Prag / I. Androutsopoulos / N. de Freitas, Machine Learning for Ancient Languages: A Survey, Computational Linguistics 49 (2023), 703–747.

Zhang et al. (2021): C. Zhang / P. Mayr / W. Lu / Y. Zhang, Extraction and Evaluation of Knowledge Entities from Scientific Documents, Journal of Data and Information Science 6 (2021), 1–5.

Author Contact Information[39]

Dr. Chiara Palladino
Assistant Professor
Department of Classics and Ancient History
Durham University
38 N Bailey
Durham, DH1 3EU
E-mail: chiara.palladino@durham.ac.uk

39