



# THE HERODOTOS PROJECT:

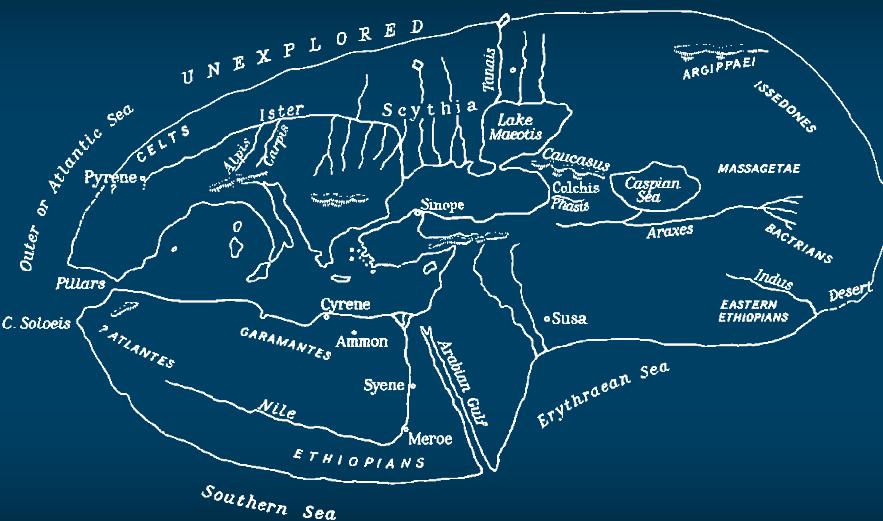
Towards an ethnohistory of the ancient world

by Christopher Brown, Marie-Catherine de Marneffe,  
Micha Elsner, and Brian D. Joseph

The ancient Greek writer Herodotus (484–425 BCE) is often referred to as the Father of History: his *Histories* tell the story of the Persian Wars or, as the ninth-century Byzantine church father Photios put it, the story of the Persian kings and an usurper. History as a discipline often focuses on events, persons, and places. But Herodotus might also be considered the Father of Anthropology, since his work is a treasure

house of a variety of information about other, non-Greek peoples of the ancient world; for this reason, he was also called Φιλοβάρβαρος (philobárbaros) ‘fond of barbarians.’ In this way, Herodotus was practicing what might be called “ethnohistory,” essentially, historical anthropology.

An ethnohistory of the ancient world as seen by the Greeks and other ancients, such as the Romans, is sorely needed. Just imagine the diversity of peoples with whom the ancient



## THE WORLD ACCORDING TO HERODOTUS | 430 B.C.

The Greeks ranged over the Mediterranean, and were in contact with indigenous peoples from Iberia in the west to India in the east, and from the Crimea and the northern shore of the Black Sea in the north to Africa in the south.



Ohio State University professors Marie-Catherine de Marneffe and Micha Elsner are two of the team members on The Herodotus Project.

Greeks came into contact! The Greeks ranged over the Mediterranean, and were in contact with indigenous peoples from Iberia in the west to India in the east, and from the Crimea and the northern shore of the Black Sea in the north to Africa in the south. Some of these peoples are well known, such as the ancient Egyptians or the Persians, or various Greek subgroups (Dorians, Ionians, etc.), while others are given only brief mention in few records, like the Libyan tribe the Asbystae (mentioned

in Herodotus 4.170) or the Sanni of Cappadocia (mentioned in Strabo XII.3.18). The Romans, with all their conquests, met with a similar range of contacts. In many cases, modern research in archaeology, anthropology, linguistics, and even genetics has added to our knowledge of these peoples. Yet no comprehensive catalogue exists of these ancient peoples, either in digital form or in print. Nor is there a single compendium of information about them and aspects of their lives, not just information known to the ancients (for instance, as seen in Herodotus's *Histories* about their habits, their language, and such) but also known through more modern scholarship (e.g., about their history, their social organization, their contacts with others, their monuments, their art, their religion, etc.).

## THE HERODOTUS PROJECT

The Herodotus Project is a digital humanities ethnohistory project which aims to fill this gap. Our goal is the compilation of all available information about these ancient peoples. It is the work of a team of Ohio State University (OSU) researchers: Brown and Joseph are specialists in Greek and Latin, and de Marneffe and Elsner are experts in computational linguistics. Because the project begins with the attestations of these peoples in Greek and Latin literary sources, we are developing a computational tool to automatically find the names of these groups in Greek and Latin texts. We are working with several undergraduate

... Herodotus might also be considered the Father of Anthropology, since his work is a treasure house of a variety of information about other, non-Greek peoples of the ancient world. . . .



and graduate students to manually annotate ancient texts in such a way that a computer algorithm can learn recognition patterns from these annotations and successfully identify group names in unannotated text; at the same time, we are adding annotations for names of persons and places so that those may be found by the algorithm as well. The task of automatically identifying such names is called Named Entity Recognition (NER).

---

## The NER aspect of the Herodotus Project was funded by a major grant from the National Endowment for the Humanities.

---

The NER aspect of the Herodotus Project was funded by a major grant from the National Endowment for the Humanities. The team has annotated texts from Caesar (*The Gallic Wars*), Ovid (*The Art of Love*), and both Plinys (Pliny the Elder's *Natural History* and Pliny the Younger's



**ABOVE:** The annotation team keeps growing, with new members like classics graduate student William Little and undergraduate computational assistant Ben Allen joining the project. **TOP:** The annotation team includes (clockwise from bottom left) Andrew Kessler, Hannah Young, Brian Joseph, Colleen Kron, Matias Grioni, Christopher Brown, Jimmy Wolfe, and Petra Ajaka.

---

epistles) with a very high rate of accuracy (over 90 percent) in identifying personal, place, and group names. We are currently perfecting the Latin NER system (originally developed by our first Graduate Research Associate Alex Erdmann) and are moving on next to develop a parallel NER for Ancient Greek, beginning with the works of Herodotus and the Roman-era Greek geographer Strabo (*Geography*). This textual annotation will also be useful for other scholars and will be made



Graduate Research Associate Alex Erdmann developed the original Latin NER system that is serving as a model for the NER for Ancient Greek.

available on public databases. We hope ultimately to extend the NER algorithm to texts in other ancient languages such as Old Church Slavonic. Computationally, the project presents interesting challenges to digital scientists, since it involves data sets that are relatively small and for which there are relatively few existing resources to learn from.

You can visit go.osu.edu/herodotos to see prototypes of some of the informational pages on ancient peoples and follow links to the various publications and presentations the team has produced. ■

#### ABOUT THE AUTHORS

**Christopher Brown, PhD**, is Lecturer in Classics at OSU, where he is responsible for the Modern Greek language program and lectures on Greek history and culture. His current projects include “Dialectics East and West,” an ongoing working group in philosophy at the Humanities Institute of OSU.

**Marie-Catherine de Marneffe, PhD**, is an assistant professor in Linguistics at OSU. She has contributed to defining the Stanford Dependencies and the Universal Dependencies, and she serves as a member of the North American Association for Computational Linguistics board and the Computational Linguistics and Lingvisticae Investigationes editorial boards.

**Micha Elsner, PhD**, is an associate professor in the Department of Linguistics at OSU. He is a computational linguist, researching how babies learn language and teaching computers to do the same.

**Brian D. Joseph, PhD**, is a Distinguished University Professor of Linguistics and Kenneth E. Naylor Professor of South Slavic Linguistics at The Ohio State University. He is currently vice president/president-elect of the Linguistic Society of America and is co-editor of the Journal of Greek Linguistics.

Ohio Humanities proudly supports cultural programming across the state.



Discover more at [OhioHumanities.org](http://OhioHumanities.org).



NATIONAL ENDOWMENT FOR THE

Humanities

