1 Introduction

1.1 Languages of Central Asia, Past and Present

South-eastern Central Asia – the region extending roughly from southern Afghanistan and north-west Pakistan to Samarkand and the Zeravshan valley in central Uzbekistan – is highly multilingual in the present day and was equally multilingual in antiquity (Figures 1 and 2). Today, the dominant language families in Central Asia are Turkic (e.g. Uzbek, Kazakh, Turkmen, Kyrgyz) and Iranian (e.g Tajik, Pamir languages, Dari, Pashto). At the south-eastern fringes of the region, we also find Dardic languages, members of the Indo-Aryan family, and Nuristani languages, members of a separate branch of the Indo-Iranian family from either Iranian or Indo-Aryan (Degener 2003). Some of these languages have millions of speakers and some are spoken only in a single valley. Even in the former Soviet Central Asian Republics, notionally based around a common Uzbek, Tajik, Turkmen or Kazakh identity, communities are linguistically diverse and individual multilingualism is common (see the studies in Ahn and Smagulova 2016). Other languages, in particular Russian and more recently English, act as pan-regional lingua francas.

The historical depth of some of these languages in the region is greater than others. Turkic languages are not attested in south-western Central Asia until the second half of the first millennium CE: the earliest written attestations of Turkic languages are found in Mongolia in the eighth century (Johanson 2022). Speakers of Iranian languages seem, from linguistic and archaeological evidence, to have moved into southern Central Asia by the second millennium BCE (see the extensive discussion of the evidence in Parpola 2003).

The purpose of this Element is to provide an introduction to the linguistic landscape of south-eastern Central Asia in the period from around the fifth century BCE to the second century CE. The languages attested by preserved texts at this period belong to the Indo-European (Iranian, Indo-Aryan, Greek) and Semitic (Aramaic) families, with one linguistic isolate, Elamite – a language with no proven genetic relationship to any other language (Tavernier 2018). This shows a very different pattern to what we might infer from evidence of historical language family movements, and indeed to the linguistic situation in the present day. The preserved textual evidence, in fact, preserves minority languages that are relative latecomers, historically speaking, much better than it does the dominant local spoken languages (Section 1.3). Aramaic and Elamite, it seems probable, arrived in the region with the Achaemenid Persian administration in the sixth

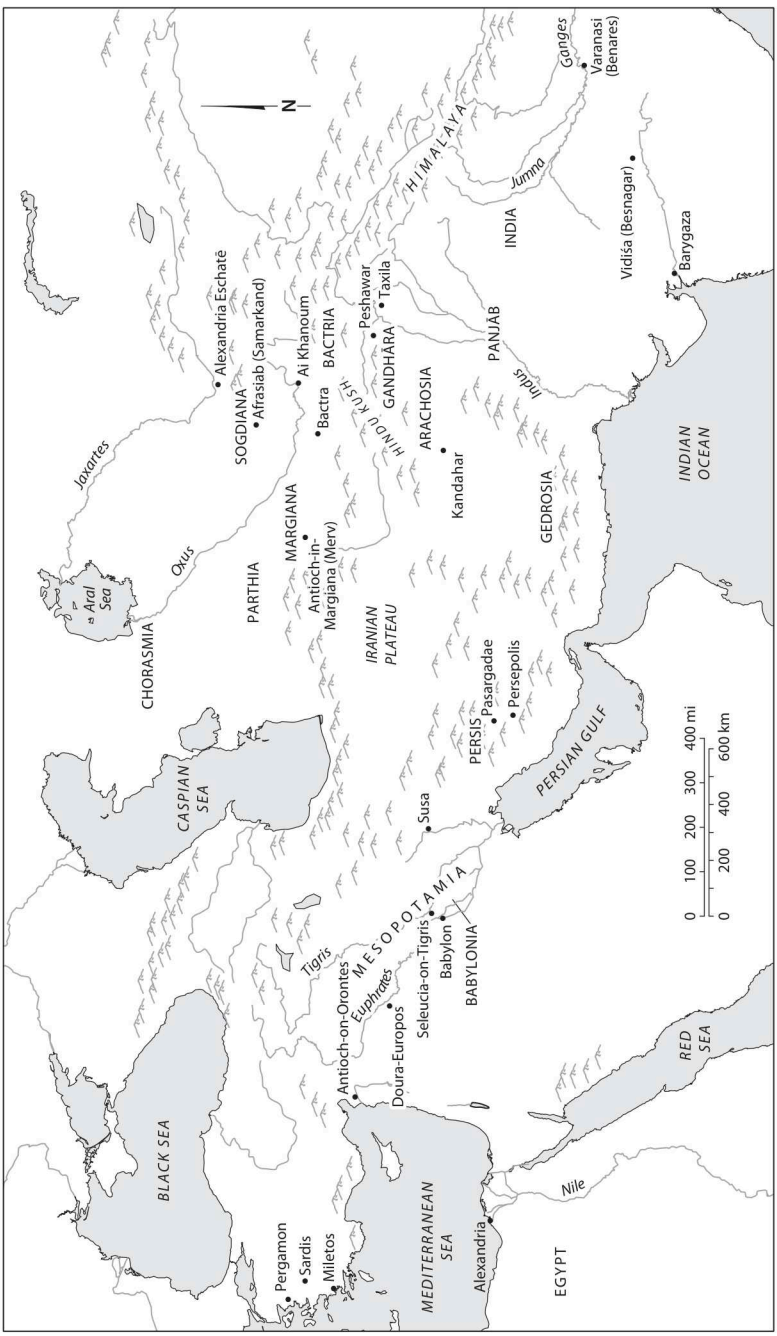

Figure 1 Map of the Hellenistic world.

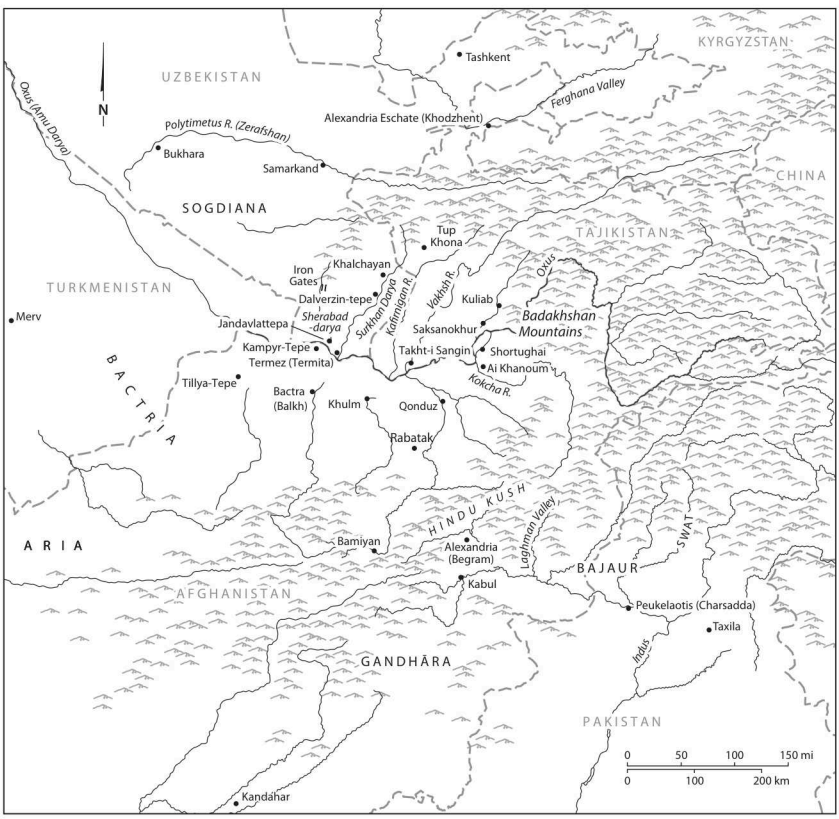

Figure 2 Map of ancient Afghanistan and surrounding regions.

century BCE. Greek is later still, with the first evidence of Greek speakers in any significant numbers in Central Asia in the late fourth century BCE. Our written sources provide only indirect evidence for spoken languages. Undocumented Iranian languages can be accessed through loanwords and linguistic interference in Aramaic, Greek and Prākrit texts. In the latter part of the period under discussion, we finally begin to find these Iranian languages attested in written form.

1.2 The Historical Context

This Element is concerned with the periods of Achaemenid, Hellenistic and Mauryan rule in the provinces of Bactria (northern Afghanistan, southern Uzbekistan and Tajikistan) and Arachosia (southern Afghanistan), with adjacent regions of Sogdiana (central Uzbekistan) and Gandhāra (eastern Afghanistan and north-western Pakistan). A fuller historical introduction

to this period, including full discussion of the archaeological and written evidence, can be found in the chapters in Mairs 2021, with a concise survey in Mairs forthcoming. These 'Upper Satrapies' were brought into the Achaemenid empire in the sixth century BCE. After the conquests of Alexander, Bactria became the centre of a Hellenistic successor state, and Arachosia and Gandhāra passed under the rule of the Mauryan empire, under the terms of a Seleukid-Maurya treaty, towards the very end of the fourth century. Around the turn of the third–second centuries BCE the Graeco-Bactrian kingdom began to expand south of the Hindu Kush, into Arachosia and Gandhāra. 'Indo-Greek' kingdoms survived in parts of this region until around the turn of the common era. The Graeco-Bactrian kingdom itself fell to a combination of nomadic invasions, war with Parthia and internal dynastic strife in the mid second century BCE. The successors to the Graeco-Bactrian kingdom were the Kushans, a dynasty who originated from the Yuezhi nomadic confederacy, from further to the north-east in Central Asia. How language and script use were shaped by these political events will be considered in greater depth in the following discussion.

1.3 Language and Script

It is important to de-couple language from ethnic identity in Central Asia. Just because someone used the Greek language to write a document, for example, does not mean that they were of Greek descent or identified on a personal level as Greek. Naming practices, too, are an imprecise metric of ethnicity, and may vary depending on context. These factors are well known in the papyrological evidence from Egypt, which shows that in the Hellenistic world people used and moved between languages in ways every bit as complex as in better documented modern societies.

As I have already noted, the languages that survive in written form in Achaemenid and Hellenistic Central Asia are overwhelmingly imperial languages: Aramaic and Elamite from the Achaemenid empire, Greek from the Hellenistic states and Prākrit from the Mauryan empire. Rather than assuming that these were also the dominant *spoken* languages in the region in this period, it is better to posit a degree of diglossia, with stable coexistence between local spoken languages (among which Iranian languages were probably dominant) and the default written registers. In the Kushan period, as shall be discussed in Section 4, Iranian languages began to be written down, using derivations of the Aramaic and Greek scripts.

Script and language must also therefore be de-coupled from one another. On several occasions in Central Asia, scripts were adapted to write languages other than the ones for which they were designed: Aramaic to write Prākrit and an Iranian language, and Greek to write Bactrian (the Iranian language of the Kushans). This shows that in Central Asia, as well as adhering to historical, imperial written registers, people were also driven to think creatively about adapting these registers to new, local needs.

Overall, the written evidence from Central Asia provides only snapshots of language use. These snapshots are determined by chance survival of evidence and by the locations in which literacy happened to have existed and been deployed in antiquity. It is nevertheless possible, and desirable, to try to bring these together into a panorama, one that may be enhanced by new discoveries in the future. I have not included every text in my survey, but have as far as possible given detailed examples from primary evidence to illustrate the linguistic points that I make.

2 Aramaic, Elamite and Iranian Languages in the Achaemenid East

2.1 The Achaemenid Administration

Bactria and Sogdiana were brought into the Achaemenid empire in the 540s BCE by Cyrus the Great (Herodotos I 177; Briant 2002, 38–40). Arachosia may also have been annexed at this time. Conquests in India followed in the 530s, and continued under Darius I in the 510s. In his trilingual inscription at Behistun, Darius claimed rule over extensive territories in Central Asia and as far as the Indus Valley (on the archaeological evidence for these, see Wu 2021 and Petrie 2021):

> Proclaims Darius, the king: These (are) the countries which fell to my lot; by the favour of Auramazda I was their king: Persia, Elam, Babylonia, Assyria, Arabia, Egypt, (the people) who (dwell) by the sea, Lydia, Ionia, Media, Armenia, Cappadocia, Parthia, Drangiana, Aria, Chorasmia, Bactria, Sogdiana, Gandāra, Scythia, Sattagydia, Arachosia, (and) Maka, all round twenty-three countries.
>
> (Trans. Schmitt 1991, 49 = Column 1, lines 13–17 of the Old Persian text.)

There are no Achaemenid royal inscriptions known from Central Asia, but we do have administrative documents from Bactria and Arachosia. These date to the fourth century BCE and constitute the earliest surviving written texts from the region. To understand how these texts came to be

written in the scripts and languages they were, we first need to consider the operation of the Achaemenid administration across the empire as a whole.

The Achaemenid empire at its greatest extent included speakers of a dizzying array of languages, from Greek and Egyptian in the west, to Iranian and Indo-Aryan languages in the east. Although the language of the ruling dynasty was Old Persian, this was only a minority language within the empire as a whole. The practical problems of running a multilingual empire were surmounted by the use of multilingual personnel, including translators, and the use of Aramaic as a common administrative language (Tavernier 2008). The Central Asian evidence shows that Elamite, too, served as a common language of record-keeping well beyond its home region in the southern Zagros.

The most extensive contemporary written evidence for the Achaemenid administration comes from Persepolis and from Egypt. We are, as ever with ancient documents, at the mercy of the gods of chance and of climate in terms of what has come down to us. Documents on perishable materials survive better from Egypt than anywhere else in the ancient world because of the dry climate and hundreds of years of modern excavation, looting and conservation. In other regions, clay tablets survive much better than texts on other media. This is true of Persepolis, where survival of the largest corpus, the Persepolis Fortification Archive, was also aided by the collapse of the building in which they were stored in antiquity.

The Persepolis Treasury and Fortification Archives were recovered from excavations during the 1930s. The texts of the Fortification Archive cover the period 509–493 BCE. There are tens of thousands of individual pieces, with the majority in Elamite in the cuneiform script and a smaller number in Aramaic. There are single tablets in a handful of other languages, including Greek and Old Persian. The content of the texts relates to the management of commodities: receipt, transfer, taxation, redistribution and so forth. Full edition and publication of the texts is ongoing (https://isac.uchicago.edu/research/projects/persepolis-fortification-archive, accessed 5 December 2025; see also Briant et al. 2008 for state of the question as of 2008). The Treasury Archive is smaller, consisting of several hundred pieces, almost all in Elamite, mostly dating to 492–57 BCE. Payments in silver are the principal topic. Because of accidents of preservation, we cannot say what, if any, administrative archive on perishable materials such as animal skin or papyrus, probably principally in Aramaic, once existed.

The preserved administrative documents from Egypt are rather different. These are a set of Aramaic documents written on skin and papyrus relating to the affairs of Aršāma, satrap of Egypt in the fifth century BCE (Tuplin and Ma 2020). Rather than documenting the management of commodities, they mostly contain instructions from Aršāma (who at the time of writing was in Babylonia or Susiana) to a variety of his subordinates (in Egypt). On the basis of this, one might propose a functional division of labour between Aramaic and Elamite in the Achaemenid administration: with Aramaic used for correspondence and long-form texts, and Elamite for records of a more practical, concise and economic nature. We can test this idea further when we come to consider the material from Central Asia.

2.2 Achaemenid Aramaic and Elamite

The forms of the Aramaic and Elamite languages used in the Achaemenid administration are slightly different from earlier standards. Aramaic is a Semitic language, which was widely spoken and written across the Middle East in antiquity. It was written in an abjad, a script in which only consonants and long vowels were routinely marked. Achaemenid official Aramaic 'bears the characteristic marks of an imperial chancery: script, spelling, morphology, syntax, lexicon, and idiom are much more unified than in the immediately preceding stage, presumably as a result of a large-scale administrative reorganisation and unification under Darius I and Xerxes' (Gzella 2021a, 118). Iranian loanwords are common in the Aršāma letters (Tavernier 2020, 77–9). We also find occasional calques on Old Persian expressions in the Aramaic of the letters (Tavernier 2020, 84–7), implying that they were drawn up by speakers of Old Persian or came from orders originally given in Old Persian, by a native speaker such as Aršāma himself. The Iranian words used in the letters are not just specialist terminology, but include everyday terms. In an instruction from Aršāma about provisioning officials on a journey to Egypt, for example, we find:

> Give them this ration, from (one) official to (the next) official, in accordance with the route which is from province to province, until he shall reach Egypt.
>
> (Trans. Taylor in Tuplin and Ma 2020, 35.)

The word for 'route' is written *'dwn'* in Aramaic,[1] and derives from an Iranian root **advan-* (Tuplin in Tuplin and Ma 2020, 178). The other words in the sentence are Aramaic.

[1] The apostrophe-like sign is used to transliterate the Aramaic letter *alaph*.