Why dedicate your life to deciphering 4,000-year-old tax returns?

Meet the elite group of academics determined to translate scrolls written in long-lost languages – one painstaking, time-consuming syllable at a time.

Words by John Preston

One day in 1994 Richard Salomon, professor of Asian Languages and Literature at the University of Washington, received a small package in the mail. Inside were a number of blurry black and white photographs and an accompanying letter from the British Library asking if they might be of any interest.

Salomon started looking at the photos – first idly, and then with growing disbelief. "I could see pretty quickly they were the real deal." The photos showed various inscriptions that were written on a series of scrolls – scrolls of bark that the British Library had been given by an anonymous donor, who in turn, had bought them from an anonymous buyer based somewhere in Pakistan.

The inscriptions Salomon saw were written in Gandhari, a middle Indo-Aryan language closely related to Sanskrit that was in use from the third century BC to the fourth century AD. It was hardly surprising that the British Library had come straight to him. Salomon was one of the few, the very few, people in the world who could read Gandhari – or at least read some of it. "I knew the basic grammar, but there were an awful lot of words that I didn't know."

Up until then Salomon had been working on the only known example of a Gandhari manuscript ever discovered – it's also reckoned to be the oldest surviving example of an Indian text. This discovery, though, changed everything.

A few days later, Salomon flew to London to have a look for himself.

Because they're written on bark, Gandhari manuscripts are much more fragile than anything on paper, or vellum. A French archaeologist who discovered some in the 1830s found that they literally crumbled to dust as soon as he touched them. Rolled up, the manuscripts Salomon saw resembled enormous cigars. Unrolled, some of them were more than 8ft long. As he gazed at them, something strange happened. "Literally, it was as if my life flashed before my eyes." Straight away, Salomon realised that there was so much new material here he was going to be spending the rest of his career working on it. Sure enough, 20 years on, he's still hard at it. "I know a lot more now than I did, but there's still a long way to go."

A life further removed from today's torrent of tweets, Facebook posts and 24-hour news is hard to imagine. For Salomon and the small band of scholars around the world dedicated to translating ancient languages, "status updates" happen only rarely and the internet's fire hose of information is more or less irrelevant.

And while it's easy to assume there are no longer any unknown languages left in the world, that they all gave up their secrets long ago, the truth is, there are lots of them. Several are well on the way to being deciphered, but others remain out of reach.

Take Etruscan, for instance. Etruscan was the main spoken and written language of the Etruscan civilisation that held sway in Italy from 700BC to 500AD. Today, we only understand a few hundred words of it. As for counting in Etruscan, if you can make it to six you're a sho-shin for a Nobel Prize. And then there's the Elamite language, spoken in Iran almost 5,000 years ago. This has had scholars banging their heads against library walls for generations – partly because it seems to bear no resemblance to any other script.

Lots of people have heard of Linear B, the ancient Minoan script found on various tablets in the palace archives in Knossos in Crete. Linear B was eventually deciphered by the British linguist Michael Ventris, who died in a car accident in 1956, just weeks before his conclusions were published. But what of Linear A, the language used in Crete before Linear B? That's proving more of an uphill struggle.
When I ask John Younger, professor of classics at the University of Kansas, how long he’s spent trying to decipher Linear A, there’s a very long pause.

“Probably about 20 years,” he says at last.

And how far has he got?

There’s another long pause. “Well, what I always like to say is that we can read it, we just don’t know what it says.”

In fact, Younger says, he reckons he now knows about half the grammar of Linear A. “We’ve certainly made great progress in the last seven years. A lot of the Linear A manuscripts, almost all of them in fact, are to do with taxes – the palace in Crete kept very detailed records of who paid what. As well as half the grammar, we now know how the administration worked at the time, so that’s a big step forward.” So, basically, he’s trying to decipher 4,000-year-old tax returns.

“I know, I know…” he says. “It’s not exactly Jane Austen. A lot of these texts are pretty monotonous, I must admit, but every so often something changes slightly – and that’s when you think, ‘What’s going on here?’”

It is small breakthroughs such as this that keep Younger, and scholars like him, going.

“These languages are as worthy of study as, say, ancient English or classical Greek,” says Susan Whitfield, an expert on central Asian manuscripts at the British Library. “What they teach us is that we live in a world that’s always been connected. For instance, the Lindisfarne Gospels [ornate manuscripts created by monks in north-east England in the eighth century] have lapis lazuli in them – and lapis comes from Afghanistan. And when you study scrolls from a monastery in central Asia, you find they connect with traditions of monasticism in this country.” But it’s not just that. However monotonous and frustrating the task of translating ancient manuscripts might be, there is still something richly romantic about the idea of a “lost language”. They reach across history, shedding light on the way people lived all those centuries ago. By deciphering a language, you open a window into the past.

When the French linguist Jean-François Champollion translated the hieroglyphics on the Rosetta Stone in 1822, he unlocked the mysteries of Ancient Egypt. And when Ventris cracked Linear B, the earliest days of Ancient Greece sprang back to life after three millennia in the dark.

As he sits poring over Gandhari manuscripts in the University of Washington, Salomon sometimes feels an almost mystical sense of connection with the monks who wrote them more than 2,000 years ago. “I really do feel there’s a link between their desks and mine. I’ve become familiar with some of the monks through their handwriting and we’ve given them names like Big Hand and Thick Hand. It’s hard to describe, but sometimes it does feel like a very personal connection.” Over at the School of Oriental and African Studies in London Nick Sims-Williams has spent the past 20 years working on translating texts written in a language called Sogdian.

Sogdian, as you may or may not be aware, was an eastern Iranian language spoken in what’s now Tajikistan around 2,000 years ago. One of the oldest manuscripts, written in around 313AD, is mounted behind glass at the British Library. “It’s a letter sent to a man’s family in Samarkand, but it never got there,” says Sims-Williams. “There had been terrible things going on in China and the man who wrote it was obviously worried about his family. He writes things like, ‘If I never make it back to Samarkand please look after my money for the sake of my son.’

“We know the letter never arrived, so God knows what actually happened to him. As you’re reading it, you do feel that this is a real person and he’s writing about these things in much the same way as we might do now.”
But how do you begin to decipher a lost language? You hope that there's something there - a letter, a symbol or a numeral - that you can make sense of. "In Linear A there are certain signs that are basically the same as in Linear B," says Younger. "That gives you somewhere to start." You then spend an awful lot of time looking at the texts and analysing the patterns of letters or symbols. "In a way it doesn't necessarily matter what the individual words mean because you can make some pretty sophisticated interpretations of what's going on, both from the patterns and from what we know about how the society worked."

It is painstaking work. Teams of researchers will routinely argue over the identity of a single letter or whether a particular civilisation used inflexives. "The other day for instance I came across the Sogdian word for liver," says Sims-Williams. "That was quite a big moment. We had the Sogdian words for all the different fingers and toes before, but not for liver." These Eureka moments tend to be pretty few and far between, though. "Slow accretion - that's really the name of the game," says Younger.

Trying to unpick a lost language is also very solitary work. "Yeah, it's not exactly something you can have out with the family over dinner," says Younger. "But that's fine for me - I love working on puzzles and I love detective work. For instance, I couldn't sleep last night so I got up at 2am and started working on Linear A." Younger receives a steady stream of carefully thought out theories from fellow specialists. But he also has to contend with a regular influx of deeply eccentric suggestions.

"Oh yes, you get a lot of nuts," he says cheerfully. "I'm a real magnet for mad people. At the moment for instance I've got one woman telling me that Linear A is Japanese, someone else saying it's Celtic and someone else saying it's proto-Persian. But like the story about the troop of monkeys eventually typing up Shakespeare, they do occasionally send in quite plausible suggestions."

But while the path may be painstaking, solitary and pitted with lunatics, it's not as slow-paced as you might imagine. "People have this image of a scholar with crumpled handwriting taking an eternity to translate one sentence, but it's really not like that," says Sims-Williams. "I mean, I've published six books in the last six years." So, does that mean that it won't be long before we know all there is to know about Sogdian?

He gives me a stern look. "No," he says. "I wouldn't say that. In fact, our work is basically infinite because there's new material being dug up the whole time." One of the most exciting discoveries in the 20th century was made by a Taoist priest called Wang Yuanli in the Mogao Caves of Dunhuang, China. While engaged in an amateur restoration of statues and paintings in what is now known as Cave 16, Wang noticed a hidden door that opened into another cave, later named Cave 17 or the "Library Cave". Inside he found tens of thousands of documents dating from 406 to 1002AD; one of the greatest treasure troves of ancient documents ever found. They were in a variety of languages, from Chinese and Sanskrit to Sogdian and the little-known Khotanese, and covered a diverse range of subjects, including the history of Buddhism, politics, folk singing and mathematics.

Dispersed all over the world in the aftermath of the discovery, they are still being translated by dozens of teams of scholars in various universities and libraries.

Imre Galambos, a lecturer in pre-modern Chinese studies at the University of Cambridge, has been working on manuscripts written in Tangut since the late Nineties.

The official language of the Tangut kingdom in what is now north-western China, Tangut was in use for around 500 years until 1500AD. Although it looks like Chinese to the untutored eye, it bears no resemblance to it at all and is often cited as the most complicated linguistic system the human mind has ever devised.

When I meet Galambos at the British Library, he unrolls a Tangut scroll for me to have a look at it. It's a beautiful thing covered in finely drawn hieroglyphics. At intervals between the hieroglyphics are red pen strokes.

"We know what all the hieroglyphics are," says Galambos. "We've known that for a long time now." But what about the red marks, I ask? He shakes his head. "I have no idea." Never mind strange symbols - some languages defy all attempts to decipher them. Indeed, some may not be languages at all. Take Rongorongo, a series of hieroglyphics scratched on to pieces of driftwood that were found on Easter Island in the 19th century. The hieroglyphs are thought to date from around 1200, but no one has a clue what they mean. And the complexities of Rongorongo are nothing compared to those of the Rongorongo Codex. The Codex, which first surfaced in Hungary in the mid 18th century, consists of 448 pages with each page having between nine and 14 rows of symbols. Some claim the symbols bear some resemblance to Hungarian, Romanian or even Hindi. But there's another, equally plausible, theory - that the whole thing is a gigantic hoax.

Much the same could be said of the Voynich Manuscript, or "the world's most mysterious book" as it's often known. This has been around since the 16th century and is written in an alphabet that no one, including top codebreakers, has come anywhere near cracking. Even if you ignore the hoaxes, or possible hoaxes, there are still plenty of other lost languages out there, waiting to give up their secrets.

"Certainly, we've still got several undeciphered languages in the British Library," says Whitfield. "But there are new discoveries being made all the time."

In fact, thanks to recent archaeological discoveries, the study of lost languages is going through another Golden Age, or something close to it. "Admittedly, you don't often run into fellow scholars on the train," says Galambos, "but it's really escalated in the last few years. Due to the political situation in Central Asia, people have been taking much more of an interest in the area and want to know about its history."

Interest has also escalated recently in Linear A. "We've made far more progress than I ever thought we would," says John Younger, in Kansas. "As well as having half the grammar, we've got 10 words that we are pretty sure are verbs."

Does he think then that he'll have cracked it by the end of his career?

There's another lengthy pause. "Well, I'm 68 now, so maybe I've got another 20 years. You know, I'm going to be optimistic and say, 'Yes'. I can't say how exactly, but I really think we're going to get there."

Materials found in Dunhuang's Library Cave can be viewed at http://idp.bl.uk