Scholars at odds over mysterious Indus script

19:00 23 April 2009 by**Ewen Callaway**

An as yet undeciphered script found on relics from the Indus valley constitutes a genuine written language, a new mathematical analysis suggests.

The finding is the latest chapter in a bitter dispute over the interpretation of "Indus script". This is the name given to a collection of symbols found on artefacts from the Indus valley civilisation, which flourished in what is now eastern Pakistan and western India between 2500 and 1900 BC.

In 2002, a team of linguists and historians argued that the script did not represent language at all, but religious or political imagery.

Ordered or random?

From an analysis of the frequency and distribution of the script's characters, the team concluded that it showed few of the hallmarks of language. Most of the inscriptions contain fewer than five characters, few of the characters repeat, and many of the symbols occur very infrequently.

The new analysis by computer scientistRajesh Rao and his team at the University of Washington in Seattle comes to the opposite conclusion.

Rao's team assessed the script samples using what is called "conditional entropy". When aimed at language, this statistical technique comes up with a measure for the "orderedness" of words, letters or characters – from totally ordered to utterly random.
"If you look at strings that contain words, then you should see that for any particular word in the string there is going to be some amount of flexibility in choosing the next word, but they're not randomly ordered," Rao says.

Which word next?

For instance, in English text, if you find the fragment "The boy went to the", there is some flexibility in what follows. Nouns like "park" and "circus" make sense, but a verb such as "eat" does not.

Rao's team applied this analysis to Indus script, Sanskrit, an ancient south Indian language called Old Tamil, and English. They also tested the conditional entropy of the Fortran computer programming language and non-languages, including DNA and protein sequences.

Indus script characters turned out to be about as randomly ordered as the other languages. Unsurprisingly, they proved less random than DNA or protein sequences and more random than the computer language, where unambiguity is essential.

Grammatical structure

"Now we can say, based on this evidence, that they probably were literate, so the big question becomes: Can you get at the underlying grammar?" Rao says. He hopes to refine his team's technique to determine the grammatical structure of Indus script and, potentially, the language family it belongs to.

"I think we are going to need more archival data, and if we are lucky enough we might stumble on a Rosetta Stone-like artefact," Rao says.

Rao's paper has already drawn a strong response from the researchers who proposed that Indus script represents religious and political symbols, not language.

"There's zero chance the Indus valley is literate. Zero," says Steve Farmer, an independent scholar in Palo Alto, California who authored a 2004 paper with two academics with the goading title "The Collapse of the Indus Script Thesis: The myth of a literate Harappan civilization."

Simulated language

As well as comparing the conditional entropy of Indus script to that of known languages, they compared it with two simulated character sets – one totally random, one totally ordered.

Farmer and colleagues Michael Witzel of Harvard University and Richard Sproat of Oregon Health and Sciences University in Portlandcontendthat the comparison with artificially created data sets is meaningless, as are the resulting conclusions. "As they say: garbage in, garbage out," Witzel says.

Unlocking history

Farmer says that the debate over Indus script is more than academic chest thumping. If Indus script is not a language, a close analysis of its symbols could offer unique insight into the Indus Valley civilisation. Some symbols are more common in some geographical locations than others, and symbol usage seems to have changed over time.
"You suddenly have a new key for unlocking how that civilisation functioned and what its history was like," he says.

J. Mark Kenoyer, a linguist at the University of Wisconsin-Madison, says Rao's paper is worth publishing, but time will tell if the technique sheds light on the nature of Indus script. "At present they are lumping more than 700 years of writing into one data set," he says. "I am actually going to be working with them on the revised analysis, and we will see how similar or different it is from the current results."

Journal reference:Science,DOI: 10.1126/science.1170391(in press)