New techniques for studying the ancient world

Introduction

In this essay I will give an overview of some of the new tools, techniques and technologies which can be used to help study the ancient world. For each of these techniques, I will explain how they can be applied to aid research in the fields of classics, ancient history and archaeology. I will also evaluate the current state of each technique, including its advantages and disadvantages, and briefly describe the research that is ongoing in that area. My overall argument is to demonstrate that the techniques and tools covered here are not only useful, but in many cases are essential tools without which there would be aspects of the ancient world that would be impossible to study effectively.

It should be noted that most of these tools are aids to current research rather than wholesale replacements for existing methods. Even when they open up whole new ways of looking at the ancient world, they are still intended to work within the existing structure as opposed to replacing it entirely.

Computer-assisted translations

Computer-assisted translations (sometimes referred to as automated translations) are a method whereby computers are used to take away some or all of the burden of translating both ancient and modern texts. Such translations can be split into two basic types: single world lookups as an aid to manual translation and automated translations of whole passages of texts, both of which I shall look at in detail here.

Single world lookups as an aid to manual translation

At the most limited level, this technique is merely an electronic dictionary, i.e. a simple mapping from a single word to its translation. However, it can also include more detailed information about the morphology of a word, such as its number and how it declines (in the case of a noun).

The main advantage of this technique is that it is much faster than using a paper dictionary. Tied into this is the fact that an electronic format can give significantly more detail than its paper equivalent, which means that,

for example, every form of a verb can be included as opposed to just its four principal parts. Examples of tools which can produce such information include the Perseus Digital Library[1] and Words[2], and a sample output from the latter program is shown below:

```
am.at           V      1 1 PRES ACTIVE  IND 3 S X
amo, amare, amavi, amatus
love, like; fall in love with; be fond of;
have a tendency to;
```

As we can see from this output, we are provided with additional information such as the tense (PRES ACTIVE IND – present active indicative) and number (3 S – 3rd person singular).

However, the significant disadvantage of this form of computer-assisted translation is no use whatsoever if the person using the tool has no knowledge of the grammar of the language being translated and a least a small amount of vocabulary, although this is more a prerequisite of the functionality provided rather than a flaw in the method itself.

The current state of single word lookups is that this is largely a solved problem, so there is little research taking place in this specific area at the present time. Most ongoing work simply involves expanding the existing dictionaries, correcting any mistakes and perhaps improving definitions.

Automated translations of whole passages or texts

Given that the task of automating the translation of individual words is more or less a solved problem, the next logical step would to be to attempt to automate the translation of whole passages or texts from one language to another. Ideally we would like to be able to feed in a text written in a foreign language (ancient or modern) and have the same text returned to us but translated into English (or whatever happens to be our first language). Whilst this may sound like a fantasy situation, several free and fee-based services already exist to perform this task, with varying degrees of accuracy. Some tools, given a sufficiently large corpus of text, can even automatically determine the language used and therefore do not have to prompt the user for this information.

[1] http://www.perseus.tufts.edu/

[2] http://users.erols.com/whitaker/words.htm

As with single word lookups, the main advantage of automating the translation of whole passages or texts is that the process is significantly faster than manual translation, and this difference becomes more acute as the size of the text increases. In addition to this, most programs are aware of more languages than most people could ever possibly hope to learn. The ability to translate modern languages, as well as ancient ones, is also useful given that there is no standard language for publishing journal articles, and it may well be the case that a piece of work which is particularly important to your research is only available in a language which you are not familiar with.

A major problem with automating the translation of entire texts is that the quality of the translation varies drastically from one implementation to another. An automated translation can result in anything from a text which is readable and useful starting starting point all the way down to complete nonsense. Not only is this a problem because , but the inherent uncertainty about the accuracy of the translation means that we do not know whether to trust it or not – it would be much more useful if we *knew* that a program would give a translation that was an accurate representation or a poor one.

A short example of the translation produced by a computer program, Google Translate[3], is shown below:[4]

French: *Des millions d'utilisateurs consultent chaque jour des annuaires pour chercher des commerces et des services de proximit. Faites en sorte qu'ils vous trouvent. Vous pouvez młme crer des bons de rduction pour remercier vos clients et en attirer de nouveaux. L'ensemble du service est gratuit.*

English: Million users consults each day of the directories to seek trade and services of proximity. Made so that they find you. You can even create reduction vouchers to thank your customers and to attract the new ones. The whole of the service is free.

As we can see, the automatic translation is not perfect – it appears to have relied on word order to an extent and has missed out some implied words such as 'of' and has mistranslated some of the words. The translation is also a bit stilted and formal, but it is still possible to get a good idea of what the French text says.

Having criticised the current state of automated translations of entire texts,

[3]http://translate.google.com/

[4]French-English translation of *http://www.google.fr/services/*

it should be noted that this is an open area of research and there are numerous academics and commercial organisations working in this field. The whole area of natural language processing, which is a core aspect of translation, is particularly active in terms of ongoing research. Automated translation can also involve the process of machine learning, in which a human translator can correct mistakes made by a machine, which will mean, in theory, that it should not make the same mistakes again. This is also an area with a significant amount of interest and ongoing research, including at the University of Manchester.[5]

In addition, even if an automated translation is not accurate enough for academic purposes yet, it may still give a sufficiently close translation to enable someone to decide whether or not it is worth getting the whole text translated properly by a human expert – a time consuming process which ideally should only be performed if the text is likely to be of some value.

Finally, the accuracy of automated translations is continuing to improve over time, so I would suggest that, whilst the current state of translations is not sufficiently accurate to warrant using the results in any serious academic work, it is perhaps an area to watch in the future.

Geographical Information Systems

Geographical Information Systems (GIS) is a method for capturing and storing geographical data, which can later be analysed in a variety of different ways. The data collected can include digital satellite images, coordinates extracted from global positioning systems, rainfall patterns and other spatial information. The first true operational GIS was the Canadian Geographical Information Systems (CGIS), which was used to determine land capability for rural areas of Canada, including land-use management and the monitoring of resources. As well as processing huge amounts of data on a national scale, the system also allowed users to focus on specific areas within the country so that resources could be managed more effectively on a local basis.

The three main advantages of GIS, in my opinion, are preservation, manipulation and varying viewpoints. Digital representations of archaeological sites allow us to view this information without actually visiting the sites

[5] http://www.cs.manchester.ac.uk/ai/research/index.php#machinelearning

themselves (naturally an initial visit is required in order to capture the information), thus preserving them as much as possible. Given the fact that we can never physically replace these sites once they have gone, and that many of them are already in a state of decay, any technology which can allow us to preserve sites whilst retaining the ability to analyse them in detail is extremely important.

Once data about a site is stored in GIS, it also possible to manipulate it without damaging the original material, something which is not possible with physical sites. In particular, GIS allows us to alter several variables and see what effects these changes would have on a given site. For example, with GIS it is possible to alter the rainfall patterns to see what might happen if an area experienced adverse weather conditions, such as extreme droughts or heavy flooding. We can also attempt to reconstruct weather and terrain conditions described in ancient texts to see what effects these conditions might have had and whether they correlate with the author's description or are perhaps instead being exaggurated and used as an excuse, for example, of the failure of a military commander in battle.

GIS also allows us to look at a site from various viewpoints, something which we could not do easily or cost-effectively otherwise. For example, when making a physical visit to a site on foot, we can only see things from the ground level, which is useful in itself because that is also how the inhabitants of a site would originally have experienced it, but it does not always give us the full picture. Often looking at a site from above is particularly useful – what appeared to be a meaningless ditch from ground level can transform into the perimeter of a Roman camp when viewed from above.

However, GIS is not without its disadvantages. A major difficulty is the time and cost involved in capturing the data in the first place. Although the information might be used hundreds or thousands of times afterwards, there is still a requirement for someone to actually go to a site and perform an extensive and methodical data capture process which, depending on the level of detail required, can be a time consuming and expensive process. Also, when this data is collected, there is no automated method for ensuring that it is sufficiently accurate, other than taking multiple readings (which of course costs time and money) and comparing them to one another. Any mistakes made in the initial data collection process may well filter through to any work based on the analysis of that material.

In terms of future development, GIS is a relatively stable technology so

it is unlikely that we will see any major changes in the next few years. Most developments are likely to be in the process of standardising specifications for GIS equipment, and reducing the size and cost of the available equipment so that they can become available to a wider audience. Another potential area for development is adding the concept of the dimension of time to GIS so that changes over a given period can be recorded, instead of merely viewing a site as a static event.

X-ray fluorescence

In the Middle Ages, parchment was often considered to be an expensive material which was difficult to get hold of. As a result, many pieces of parchment were reused by scraping away the original text and laying down new writings and paintings on top of it. This overwriting meant that the original text could no longer be seen with the naked eye and was effectively lost to us. The exact procedures involved in applying the x-ray fluorescence technique are too complicated to explain here, but the basic principle is that the parchment is bombarded with gamma rays which the iron contained within the ink of the original text will react to. The overall effect of this process is that the original text appears to glow slightly, revealing some or all of the overwritten words.

In some cases the only remaining copies of texts are hidden in this way, for example the original Greek text of *On Floating Bodies* by Archimedes was thought to be lost and was only rediscovered in 2006 as a result of applying this technique. As well as uncovering hidden texts, it can also be used to decipher stone inscriptions (which we often only have one copy of) which have been worn down over time and are no longer visible with the naked eye.

The obvious major advantage of this technique is that it allows us to view texts which would otherwise be unavailable to us. Without it, we would not be able to take account of

Another advantage which makes this technique so useful is that its application is non-destructive, thus it can be safely used on rare and valuable manuscripts that are already in a state of detioration. The fact that this technique does not damage the original parchment in any way means that it can be used on any document that may contain a hidden text, . This is in contrast to techniques which may damage or contaminate the material

during their application, such as radiocarbon dating, where it is important to decide whether or not the application of the technique will yield any useful data beforehand and whether that information is worth more than the potential damage or contamination caused by the application of the technique. Estimating the potential value of data gained from the use of such a tool is extremely difficult, so having a technology which bypasses this issue by not damaging the original material in the first place is a significant advantage.

As with any technique, x-ray fluorescence has its downsides. A significant problem is the speed at which the process takes place – it can often take twelve hours or more to scan a single page, and this does not include the time required for the painstaking process of manipulating and enhancing the scan in order to take it from a blurred digital image into an actual text that scholars can work with. After all this effort, there is no guarantee that the end result will be of use – it could be the case that, even after cleaning up the image, the text is still too distorted to be of much use, or that it turns out to be merely a copy of a text which is already extant.

The problem of the speed at which this technique operates is compounded by the cost of the equipment required to apply it. As a direct result of this expense, only a very small number of institutions can justify purchasing the equipment and so they must rely on the generosity of those organisations which do have access to such resources. This limited supply of equipment, combined with the time taken to produce any results, means that at present the application of this technique has to be prioritised to those documents thought to contain the most important material, rather than being applied to any document which might reveal some useful information.

Points raised in discussion

A number of points were raised in the discussion period following the presentation, on a wide range of issues and topics. There is insufficient space to discuss all of them in depth here, so I have chosen to focus on three particular issues which I felt were the most relevant to the topic of the presentation and which reflected the broad scope of the questions in general.

The first point that was raised was a question about whether the applica-

tion of the x-ray fluorescence technique can be used to reveal texts composed with vegetable inks. Unfortunately, I have been unable to find any material which exactly answers this specific question, but the majority of the texts that I came across seemed to imply that, as the technique works with carbon-based inks, it should work on parchments containing texts composed with vegetable inks.

The second question concerned whether any of these techniques could shed any new light on classical sites which have already been studied in depth. The general answer, in my opinion, depends on a number of factors, including what you are looking at and hoping to find, as well as the site itself. Most sites would probably benefit from having survey information entered into GIS, if nothing else this will make the data easier to analyse and manipulate – although this has probably already been done for most major sites. The application of techniques such as x-ray fluorescence relies on appropriate material (e.g. stone inscriptions) being available, which may not be the case for all sites, but where such material does exist I would suggest that there is a possibility, sufficiently strong to warrant investigation, that new information could be obtained by considering the use of these new tools.

The final issue that was raised was how to deal with material that has not been digitalised. Obviously any data which is not already in a digital format cannot have some of the techniques described in this essay applied to it, and that in itself is a barrier to their use. However, scanning in documents is a relatively easy task, especially compared with the time and error factors involved with manually typing in all the text from a given work. The equipment required for scanning in material is cheap and readily available, well within the reach of any departmental budget (and even the individual researcher in some cases). The only issues with digitalising large amounts of material is the time required for someone to invoke the process on each individual page of a work and to correct any errors that occur when the scanned image is transformed into digital text. The former problem can be mitigated by providing the material in a loose-leaf format which can be automatically scanned page by page without the need for manual intervention, in a similar way to photocopying an unbound article. The latter issue, whilst still occasionally problematic when dealing with some texts, is mostly overcome by the fact that Optical Character Recognition (OCR) software is now sufficiently advanced to transform a scanned image into text with a high degree of accuracy, requiring only a small number of manual corrections.

Conclusion

Having looked in detail at several of the new techniques available for studying the ancient world, my overall conclusion is that these tools can be of great help in specific areas, without requiring a change in existing research methods. Whilst none of the techniques described here are essential to studying the ancient world (although some, particularly GIS, have become standard research tools that everyone is aware of), most of them (with perhaps the exception of automated translation in its current state) can yield valuable information and insights which would not be possible ot gain through existing methods.

Bibliography

Fildes, J. (2006), 'X-rays reveal archimedes secrets', BBC News online. Accessed 7th March 2007.
URL: *http://news.bbc.co.uk/1/hi/sci/tech/5235894.stm*

Lock, G. (2003), *Using computers in archaeology : towards virtual pasts*, Routledge, London.

Lock, G. & Brown, K., eds (2000), *On the theory and practice of archaeological computing*, number 51 *in* 'Oxford University committee for archaeology monograph', Oxford University Committee for Archaeology, Oxford.

Martlew, R., ed. (1984), *Information systems in archaeology*, New standard archaeology, Alan Sutton, Gloucester.

Reilly, P. & Rahtz, S., eds (1992), *Archaeology and the information age: a global perspective*, number 21 *in* 'One world archaeology', Routledge, London.

Richards, J. & Robinson, D., eds (2000), *Digital archives from excavation and fieldwork : a guide to good practice*, AHDS guides to good practice, 2nd edn, Oxbow, Oxford.