

Reading the Readers: Modelling Complex Humanities Processes to Build Cognitive Systems

Melissa Terras

University College London, UK

Abstract

The ink and stylus tablets discovered at the Roman Fort of Vindolanda are a unique resource for scholars of ancient history. However, the stylus tablets have proved particularly difficult to read. This paper describes the initial stages in the development of a computer system designed to aid historians in the reading of the stylus tablets. A detailed investigation was undertaken, using Knowledge Elicitation techniques borrowed from Artificial Intelligence, Cognitive Psychology, and Computational Linguistics, to elicit the processes experts use whilst reading an ancient text. The resulting model was used as the basis of a computer architecture to construct a system which takes in images of the tablets and outputs plausible interpretations of the documents. It is demonstrated that using Knowledge Elicitation techniques can further the understanding of complex processes in the humanities, and that these techniques can provide an underlying structure for the basis of a computer system that replicates that process. As such it provides significant insight into how experts work in the humanities, whilst providing the means to develop tools to assist them in their complex task.

1 Introduction

The ink and stylus texts from Vindolanda are an unparalleled source of information regarding the Roman Army and Roman occupation of Britain for historians, linguists, palaeographers, and archaeologists. The visibility and legibility of the handwriting on the ink texts can be improved through the use of infrared photography. However, due to their physical state, the stylus tablets (one of the forms of official documentation of the Roman Army) have proved almost impossible to read.

This paper describes the techniques used to develop a system to aid historians in reading the stylus texts and, in the process, developing what appears to be the first computer program to aid experts in reading an ancient document. Before designing and building any tools to aid

Correspondence:

Melissa Terras,
School of Library,
Archive and Information Studies,
Henry Morley Building,
University College London,
Gower Street,
London WC1E 6BT,
UK.

E-mail:

m.terras@ucl.ac.uk

papyrologists in the reading of texts, it is a necessary requirement to ask first: just what does a papyrologist do when trying to read and understand an ancient text? Although the readings generated from ancient documents provide one of the major primary information sources for classicists, linguists, archaeologists, historians, palaeographers, and scholars from associated disciplines, surprisingly little research has been carried out regarding how an expert constructs meaning from deteriorated and damaged texts.

This paper discusses an interdisciplinary approach to modelling a complex humanities process, where techniques from artificial intelligence, cognitive psychology, knowledge elicitation, computational linguistics, and content analysis, are combined to result in a proposed model of how experts read ancient documents. This representation was subsequently used as a basis for the development of a computer system that can aid historians in the reading of the Vindolanda texts.

The system is not an 'expert system' that automatically 'reads' and provides a transcription of the texts, rather it is a papyrologist's assistant that mobilizes disparate knowledge, such as linguistic and visual clues, and uses these to speed up the process by which an expert can arrive at the most likely interpretation of a text.

Constructing an explicit model of a process is the first stage in building a computer system that replicates that process, and it is demonstrated that the analysis of complex humanities procedural tasks in this manner can result in computer systems that aim to aid experts to carry out those tasks more efficiently.

Section 2 of this paper provides background information regarding the Vindolanda texts, and Section 3 discusses related research undertaken on the stylus tablets and stresses the need for the development of a cognitive visual system to aid in the reading of the tablets. Section 4 details the Knowledge Elicitation techniques utilized in order to construct an explicit model of the processes used to read an ancient text, and Section 5 describes how this model was used as the basis of an architecture to construct a system that reads in images of the Vindolanda texts and outputs plausible interpretations of the documents, to aid the papyrologists in their task.

2 The Vindolanda Texts

The discovery of the tablets at Vindolanda, a Roman Fort built in the late 80s AD near Hadrian's Wall at modern day Chesterholm, has provided an unparalleled resource regarding the Roman occupation of northern Britain and the use and development of Latin around the turn of the first century AD. Textual sources for the period in British history from AD 90 to AD 120 are rare, and the ink and stylus tablets are a unique and extensive group of documents providing a personal, immediate, detailed record of the Roman Fort at Vindolanda from around AD 92 onwards (Bowman and Thomas, 1983, 1994, 2003; Bowman, 1997).

The ink tablets, carbon ink written on thin leaves of wood cut from the sapwood of young trees, have proved the easiest to decipher. In most cases, the faded ink can be seen clearly against the wood surface by the use of infrared photography, a technique used frequently in deciphering ancient documents (Bearman and Spiro, 1996). The majority of the 600 writing tablets that have been transcribed so far contain personal correspondence, accounts and lists, and military documents (Bowman and Thomas, 1983, 1994, 2003).

The 200 stylus tablets found at Vindolanda appear to follow the form of official documentation of the Roman Army found throughout the Empire (Turner, 1968; Fink, 1971; Renner, 1992). It is suspected that the subject and textual form of the stylus tablets will differ from the writing tablets, as similar finds indicate that stylus tablets tended to be used for documentation of a more permanent nature, such as legal papers, records of loans, marriages, contracts of work, sales of slaves, etc (Renner, 1992), although the linguistic aspects of the tablets will be similar as they are contemporaneous documents from the same source, probably written by the same scribes.

Manufactured from softwood with a recessed central surface, the hollow panel of the stylus tablets was filled with coloured beeswax. Text was recorded by incising this wax with a metal stylus, and tablets could be re-used by melting the wax to form a smooth surface. Unfortunately, in nearly all surviving stylus tablets¹ the wax has perished, leaving a recessed surface showing the scratches made by the stylus as it penetrated the wax² (see Fig. 1). In general, the small incisions are extremely difficult to decipher. Worse, the pronounced wood grain of the fir wood used to make the stylus tablets, staining and damage over the past 2,000 years, and the palimpsestic nature of the re-used tablets further complicate the problem; a skilled reader can take several weeks to transcribe one of the more legible tablets, whilst some of the texts defy reading altogether. Prior to this research, the only way for the papyrologists to detect incisions in the texts was to move the text around in a bright, low raking light in the hope that indentations would be highlighted and candidate writing strokes become apparent through the movement of shadows, although this proved frustrating, time consuming, and insufficient in the transcription of the texts.

- 1 It is suspected that around 2,000 of such tablets exist outside Egypt (Renner, 1992).
- 2 Only one stylus tablet, 836, has been found so far with its wax intact. Unfortunately this deteriorated during conservation, but a photographic record of the waxed tablet remains to compare the visible text with that on the re-used tablet.

3 Research and the Vindolanda Texts

Novel imaging techniques have been developed at the Department of Engineering Science, University of Oxford, to analyse these texts (Bowman *et al.*, 1997; Schenk, 2001; Molton *et al.*, 2003; Brady *et al.*, 2004; Pan *et al.*, 2004). However, whilst a scrutiny of the document surface using image processing techniques provides new information, it does not provide a means of sorting through this information to provide the most likely interpretation of the data. Therefore, it was necessary to develop a computer system to aid the historians in the reading and interpretation of these images themselves, in order to



Fig. 1 Stylus tablet 836, one of the most complete stylus tablets unearthed at Vindolanda. The incisions on the surface can be seen to be complex, whilst the wood grain, surface discoloration, warping, and cracking of the physical object demonstrate the difficulty papyrologists have in reading such texts

speed up the reading process. The system is designed to work in combination with these image processing algorithms and the experts themselves.

In addition, linguistic analysis of the Vindolanda ink texts was undertaken (Terras, 2002), which provided lexicostatics with word lists, letter frequency, and bigraph frequency: this provided statistics regarding the language used at Vindolanda, which were available for integration into the system described here.³ A corpus of images of the ink texts was annotated to provide data regarding the letter forms found on the Vindolanda texts (Terras and Robertson, 2004).

4 The Process of Reading an Ancient Document

In order to identify the tools that could be built to aid the papyrologists in their transcription of the Vindolanda tablets, it was first necessary to try and gain an understanding of what the papyrology process actually entails. Little investigation has been done so far to ascertain how experts read such damaged and abraded documents (Terras, 2002).⁴ Techniques borrowed from the field of Knowledge Elicitation were used to gather quantitative and qualitative information about how papyrologists work, resulting in an in-depth understanding of the ways

³ For a discussion regarding how the language contained in the stylus tablets relates to that of the ink tablets, see Terras (2002).

⁴ Aalto (1945), Youtie (1963), Youtie (1968), and Bowman and Tomlin (2004) are the only discussions found (as yet) which try to describe what the papyrology process actually consists of, with some higher level discussion available in Turner (1968).

different experts approach and reason about damaged and abraded texts. The process was resolved into defined units, with characteristics about each being documented. General procedural information was also collated. This results in a proposed model of how experts read ancient documents, which was subsequently used as a basis for the development of the computer system that replicates this process, and so aids the papyrologists in their task.

The primary questions to be asked in this study were: Is there a general process that experts use when reading ancient texts? Can this procedure be elucidated? What are the differences and similarities between individual experts' approaches to the problem? In trying to answer these questions, computational techniques were employed to interrogate and manipulate any data collected, and assist in developing a model of how experts operate in the given domain.

4.1 Knowledge elicitation

The problem with trying to discover the process that papyrologists go through whilst reading an ancient text is that experts are notoriously bad at describing what they are expert at (McGraw and Harbison-Briggs, 1989). Experts utilize and develop many skills that become automated and so they are increasingly unable to explain their behaviour, resulting in the troublesome 'knowledge engineering paradox': the more competent domain experts become, the less able they are to describe the knowledge they use to solve problems (Waterman, 1986). In addition, although knowledge acquisition and elicitation⁵ are becoming increasingly necessary for the development of computer systems, there is no consensus within the field as to the best way to proceed in undertaking such a study. The approach taken in this investigation is often referred to as the 'traditional' or 'transfer' approach to Knowledge Elicitation (McGraw and Harbison-Briggs, 1989; Waterman, 1986; Diaper, 1989), utilizing protocols that were developed to aid a knowledge engineer in interacting with a domain expert to organize and formalize extracted knowledge so that it is suitable for processing by a knowledge-based system. These techniques for the capture of knowledge, often derived from clinical psychology and qualitative research methods used in the social sciences include:

- (1) unstructured, semi-structured, and focused interviews with the expert(s);
- (2) Think Aloud Protocols (TAPs), where an expert is set a task and asked to describe their actions and thought processes, stage by stage;
- (3) sorting, where the expert is asked to express the relationship between a pre-selected set of concepts in the domain;
- (4) laddering, where the expert is asked to explain the hierarchical nature of concepts within the domain.

⁵ Knowledge acquisition is conventionally defined as the gathering of information from any source. Knowledge elicitation is the subtask of gathering knowledge from a domain expert (Shadbolt and Burton, 1990).

4.2 Knowledge Elicitation and Vindolanda

A number of steps and observations were undertaken to gain an understanding of the general process the experts utilize when approaching an ancient text, and specifically, the Vindolanda ink and stylus tablets. First, as with all knowledge acquisition tasks, the domain literature was researched. Secondly, any other associated literature was collated. Although not a direct comment on the act of reading and transcribing, the two published volumes regarding the Vindolanda ink tablets contain detailed apparatus of the individual texts (Bowman and Thomas, 1983, 1994). The standard publication format (the transcribed text, marked-up using the Leiden system, followed by the critical apparatus: variant readings, misspellings, line by line comments and explanations, and the translation of the text),⁶ is all that is presented of the process that was undertaken in the reading and understanding of the documents. This apparatus aims to cover comprehensively the difficulties, reasoning, and alternative hypotheses regarding the final transcription. As such, it is the best representation of the different types of knowledge used in the reading of texts available without carrying out further investigation (although the sequential order of the different stages in their reading are lost due to the reporting format). These texts were obtained in digital format to enable in-depth study.

Three experts were then identified who were working on the ink and stylus texts, and who were willing to take part in this investigation. (They shall be referred to as Expert A, Expert B, and Expert C,⁷ so as to spare their blushes, and not to imply any criticism of their work: this investigation is concerned with asking how an expert operates, rather than making judgements on those operations.)

A series of investigations were carried out, utilising Knowledge Elicitation techniques. The experts were observed whilst going about their tasks, and unstructured interviews were undertaken, where the experts described their domain, and the individual processes and techniques that they preferred. More structured interviews then took place, when the experts were asked to describe particular facets of their work, ladder and sort different types of information used in tasks such as the identification of letter forms, and explain the role of grammar, word lists, and external historical and archaeological resources in the reading of the documents. Hourly sessions were carried out with each expert, on average twice a month for almost a year, indicating the significant investment in time and commitment necessary from both the knowledge engineer and the experts themselves.

A series of Think Aloud Protocols (TAPs) were then undertaken (a technique adopted from experimental psychology,⁸ where the expert is urged to utter every thought that comes to mind whilst undertaking a specified task) and the experts were given structured tasks to complete. These involved being presented with various images of ink and stylus tablets they had not already read, and asked to come up with the best reading they could, in some cases at their own leisure, and in others in

6 As an example of this format, this is the published commentary of stylus tablet 836 (shown in Fig. 1), by Bowman and Tomlin (2005, p. 6). In the text presented below, letters printed in boldface are those which can be read with confidence; letters printed in ordinary type are read with some measure of conjecture; underlinings indicate traces of letters which cannot be identified with confidence:

banus bello suo salutem
(traces only)
acc__erunt **in in** uecturas
de_arios octo reliquos solues
5 **rios nouem qua__r__**
sam dari debeb__
(interlinear addition?)
em libris
dus uale

‘Albanus to his Bellus greetings . . . they have received for transport costs 8 denarii. You will pay the remaining 9 denarii . . . ought to be given (?) . . . nine pounds (?) . . . Farewell.’

Notes:

There is a trace between the first and second **I** in **bello** which might or might not be a letter. The scratches on the wood show that this overlies an earlier text.

The correct reading is almost certainly **acceperunt**.

The word at the end of the line presents particular difficulty. Of the first three letters of **solues** only the **o** is certain. There is a clear high horizontal which has to be ignored if the first letter is read as **s**. The third letter might be **p**, and there is another apparent high horizontal which is discounted. The attraction of reading the

word **solues** (from the verb **soluere** 'to pay') is obvious if the word '**denarios**' occurs twice in lines 4–5.

7 All three experts are English male academics who have been working on these texts for over 25 years. They graciously gave their time, and permission, for this study.

8 Although 'an expensive and meticulous research method that has had its share of growing pains' (Smagorinsky, 1989, p. 475), the collection of verbal data in this manner has been a procedure used in the social sciences for three-quarters of a century (Duncker, 1926). Protocol analysis has been shown to be 'a very useful addition to the repertoire of research tools . . . The data from most other tools yield little about the internal structures of cognitive processes, particularly when the tasks are complex. Think-aloud protocols, in contrast, can yield significant information about the structure of processes' (Smagorinsky, 1989, p. 465). Most cognitive studies of translation and interpreting use TAPs as the tool of choice (Danks *et al.*, 1997, p. xv), for example Kiraly (1997) effectively used TAPs to investigate how translators work.

9 <http://www1.oup.co.uk/elt/catalogue/Multimedia/WordSmithTools3.0/>

10 *Text Analysis Computing Tools*: <http://www.chass.utoronto.ca/cch/tact.html>

11 There are many different ways in which to break down a text in Content Analysis, such as word, sentence, paragraph, item, or theme (Holsti, 1969, p. 180).

the presence of the knowledge engineer. These sessions were recorded, and transcribed, resulting in 23,000 words of discussion: 16,000 words pertaining to the reading of the ink tablets and 7,000 regarding the reading of the stylus tablets. Along with the published commentaries, these texts were subject to Content Analysis techniques, and linguistic analysis (using WordSmith⁹ and TACT¹⁰), to detect any underlying structures and decision matrices. The data captured provide explicit and quantitative representation of the way the papyrologists approach damaged and abraded texts. General procedural information was also collated.

4.2.1 Content Analysis techniques

Content Analysis, a 'method of studying and analyzing written (or oral) communications in a systematic, objective, and quantitative manner' (Aiken, 1971, p. 433) is an unobtrusive, context-sensitive, empirical process in which texts are reduced and condensed into a numerical format in order to estimate some phenomenon in the context of data (Holsti, 1969; Krippendorff, 1980; Stemler, 2001). Content Analysis is the most thorough and useful way to gain an empirical insight into the structure and content of complex textual sources, and is the standard technique used in encoding and analysing the data from TAPs (Ericsson and Simon, 1993). Texts are divided into defined units¹¹ and labelled, and these units can be used as a basis for statistical analysis.

In the case of the data from Vindolanda, the *subject* of the linguistic unit was identified as the defining feature, and the text split into sections where the subject of the phrase, sentence, or sometimes paragraph, changed. A number of pilot studies were undertaken before settling on an inventive encoding scheme that comfortably encompassed the data from the apparatus, transcripts, and preliminary knowledge elicitation exercises, resolving the different types of knowledge presented into an overall framework. The final, novel encoding scheme is presented here: the very act of resolving the process into such defined units provides a modular basis on which to develop a model of the papyrology process. The TAPs, and a selection of the published commentaries, were encoded in this manner (Table 1).

4.3 Results

The data collected give a rich source for the analysis of how experts read ancient texts, and results in some interesting observations.¹² It is perhaps unsurprising that, superficially, the three experts used in this study seem to read documents differently. The individual tools and techniques they prefer differ greatly: Expert A makes most use of digital images and PhotoShop to examine the texts; Expert B favours drawing his own representation of the text as he reads; Expert C relies mostly on photographs. The three experts spent various amounts of time discussing the texts, and the word counts and timings of these discussions also varied greatly.

Table 1 The encoding scheme resolved from an analysis of the Vindolanda textual data, an original scheme devised for this project

Reading level	Thematic subject
8	Meaning or sense of document as a whole
7	Meaning or sense of a group or phrase or words
6	Meaning or sense of a word
5	Discussion of grammar
4	Identification of possible word or morphemic unit
3	Identification of sequence of characters
2	Identification of possible character
1	Discussion of features of character
0	Discussion of physical attributes of the document
-1	Archaeological or historical context ¹³

Table 2 Excerpt from Expert B's discussion of 1491 illustrating the appropriation of Reading Levels to the text, and demonstrating the flow of the discussion between different levels

Timing (secs)	Transcribed text	Reading Level	
131	Then, when we start reading it, well this looks like a very good Latin word FECIT - F E C I T.	4	Identification of Word
139	And then immediately at the end what looks like Q U A N E.	2	Identification of Characters
161	Which is a bit of a problem . . . if that is the end of the line is has to be QUA ME instead of QUAN, You wouldn't really expect to see it there.	4	Identification of Word
167	But immediately when one gets as far as this you can see that this stoke coming down is E.	1	Identification of Feature
174	So it's a descender from the line above, so we haven't got the top of N.	1	Identification of Feature
182	Before that we have got M A possibly M I N A.	2	Identification of Characters
186	Which obviously works with the Latin word DOMINA.	4	Identification of Word
190	A woman has done something.	6	Discussion of Word Meaning
202	QUA is then better, the woman is then in the perfect.	5	Discussion of Grammar

The order in which experts discussed key features of the documents and identified constituent words were also very different. For example, whilst discussing ink tablet 1491, the identification of the words in the document happened in very different orders. This can be illustrated by plotting where three words from the text, *adfectum*, *fecit*, and *qua*, were used (and therefore identified) in the course of the discussion (Fig. 2).

Although on the surface, the three individual papyrologists seem to discuss these texts in different manners, using different techniques, and in different orders, the readings that emerged from these different sessions were, for the most part, converging towards the same conclusions, and highlighting the same areas of difficulty. For example,

12 A full analysis of the data can be found in Terras (2002, chapter 2): highlights are presented here.

13 This is presented as '-1' to mark the fact that the experts are explicitly referring to other sources, and not only this document.

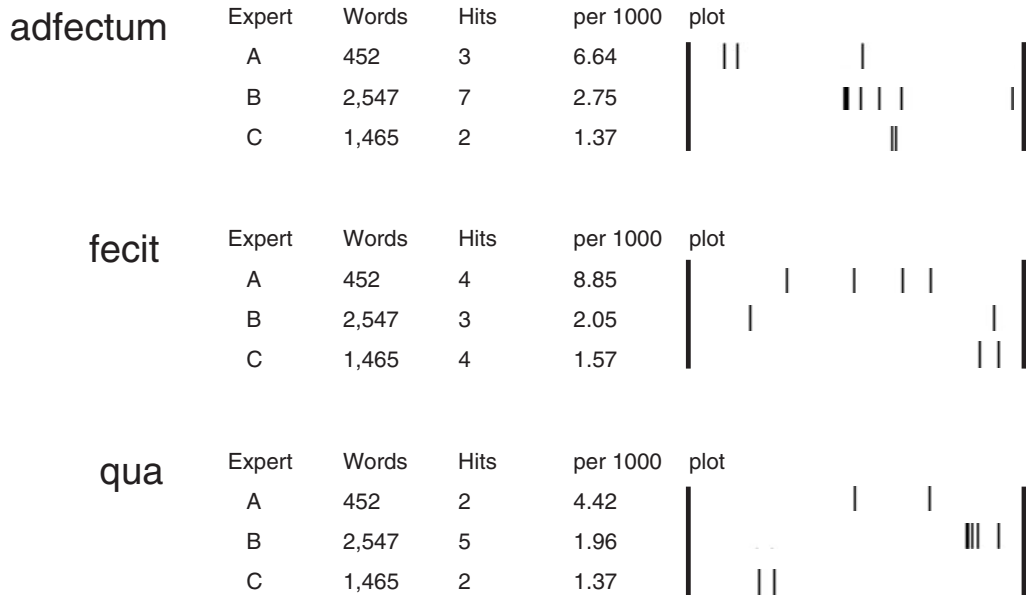


Fig. 2 Occurrence of key words of document in discussion of ink tablet 1491 by three experts. It can clearly be seen that the experts identify the words in different orders (Expert A: *adfectum* → *fecit* → *qua*; Expert B: *fecit* → *adfectum* → *qua*; Expert C: *qua* → *adfectum* → *fecit*) and at different times in the discussion

compare the three experts' readings of 1491:

And we have got something dot MAE... FECIT QUA ME
something CUNDE CONSOLARIS SIC UT MATER ET HUNC
ENIM ADFECTUM. (Expert A)

FECIT QUA ME... and then in the next line what I was reading
as FACUNDE but actually may be ECUNDE... And then this
CONSOLA word, And then another letter, And then what looks
like UT MATER... Then what I think is this sequence NUNC
ENIM ADFECTUM ANIMUS ME. (Expert B)

FECIT...QUA ME... before that we have got MA possibly
MINA... So you then pick up the individual letters
C U N D E... that could be CONSOLARIS...SICUT... And
then MATER looks good, Then this could be something from the
verb FACEO... This...reading is clear, HUNC ENIM
ADFECTUM...ANIMUS. And then you have M E and what
looks like part of a U. (Expert C)

All three experts show the same confusion surrounding the reading of the word(s) prior to FECIT, the word which may be CUNDE, and the letters around SICUT. All clearly read FECIT, had some difficulties with the sequence QUA ME, and wondered about the possible meaning of ADFECTUM whilst clearly reading the words around it. Further

analysis of the TAPs, utilizing Content Analysis techniques, revealed hidden similarities between the experts' processes, which illustrate the similarities in the way they reason about the texts.

4.3.1 *The cyclic reasoning process*

When the experts discuss the overall process they use when transcribing a text, they are very clear that it is not a process of transcribing letter by letter (as Youtie, 1963, 1966; Aalto, 1945; Turner, 1968; and Bowman and Tomlin, 2004 noted), but rather of proposing hypotheses and reasoning about these as more information comes to light:

Some people when faced with something like this will start by saying well we can identify 1,2,3,4, lines of writing here and start at the beginning of the first line and work their way through 'til the end of the last line and I suppose the idea is that you get some sort of objective view of the writing or build up of the letter, letter by letter, but I don't work that way, I never have and I don't believe that it really works very well. So what I have actually done, as I do with all these things is really go the point at which I think I can identify some of the letters in some of the words to start with and that begins to give me some sort of clue. (Expert A)

I suppose what I would do would be try and do a letter by letter transcription and start seeing if anything was making sense... this is going to be a series of interlocking hypotheses which don't necessarily resolve themselves. In a way one is doing what one often does do, which is to at this point you are sort of on the hypothesis, but you can't really be sure that it is so until you find something that kind of makes such obvious sense, that it must be right, whereas the first two lines or so do seem to work and are self contained. I don't know if you ever have tried life drawing, but its often that you draw part of the figure and it all fits beautifully. Then you find maddeningly that the foot or something is too close, and you kind of doing what kind of requires a strength of character which is to redraw the good bit, to make it fit in with the other bit, otherwise you lose the proportion, the relationship of the whole, so that it is always the problem in reading a text; how long you hold onto something that you are certain of, if it just won't fit in with anything else. (Expert B)

I could make out individual letters at first. What we're trying to do, or what I'm trying to do, is get words to make sense... not individual letters... (Expert C)

The identification of the core subjects covered in these discussions (Table 1) shows the novel scheme that can be used to illustrate the fact that discussions regarding texts vacillate between the identification of features, letters, and words, and the production of meaning regarding these components. When plotted over time it becomes obvious that

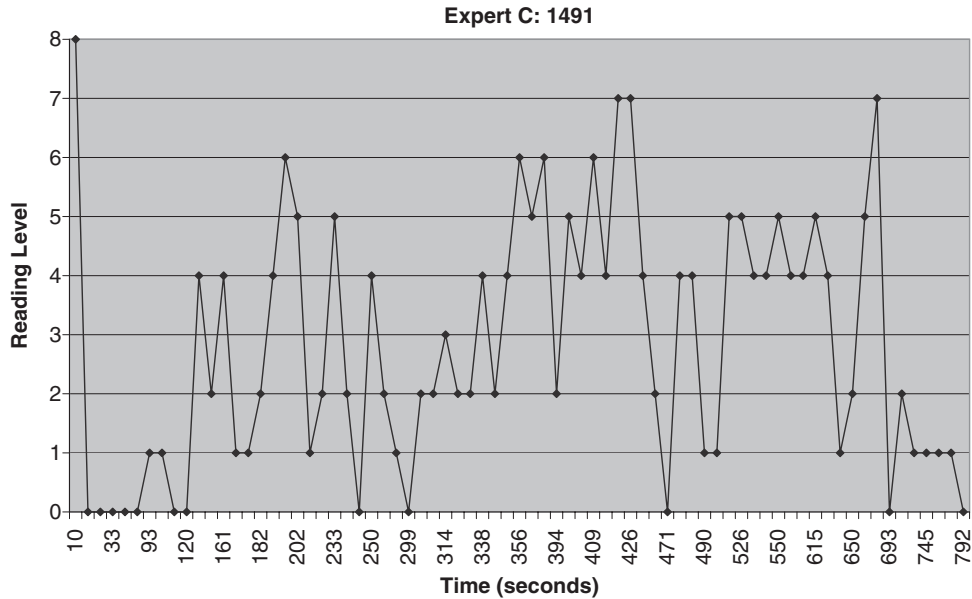


Fig. 3 Discussion of tablet 1491 by Expert C, plotted by subject matter over period of discussion

the reasoning process is far from linear, and depends on a complex cycle of interlocking elements. This can be seen in Expert A's discussion of reading ink tablet 1491. Similar graphs result when any of the discussions are plotted in this manner (Fig. 3).

Expert C begins by drawing some conclusions about the meaning of the document (level 8) before looking at its physical attributes (level 0). He then discusses what could be possible features of the text (level 1), before noting more physical attributes of the document (level 0). He then produces a word (level 4), looks at the characters within this word (level 2), and revises his initial word. Checking of the features (level 1) leads to identification of a character (level 2), the noting of a possible word (level 4) and a discussion of meaning of that word (level 6). In this manner the expert vacillates between the different levels in reading a document, until a resolution is reached regarding the sense of the document (level 8), or until he has exhausted all possibilities regarding the text. An extract from this discussion, illustrating how the Reading Levels are appropriated, is presented below.

All the experts' discussions regarding the texts followed a similar pattern, with various hypotheses concerning the identification of features, characters, and words, being checked against other information from the document, until some resolution of ambiguity was reached. Modelling this process computationally, by implementing each different level as an agent, and using techniques derived from artificial intelligence as a means to pass information between levels (see Section 5) provided a way to construct a tool to aid the papyrologists in their reading, as it enabled this cyclic process to be replicated.

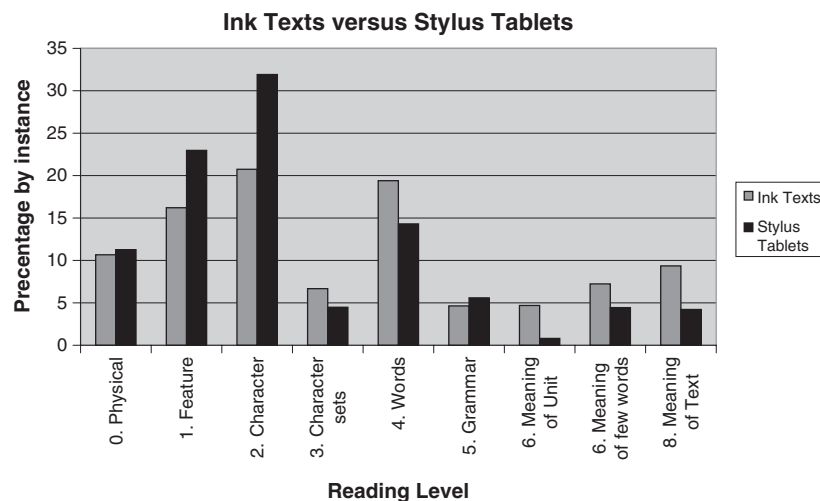


Fig. 4 Comparison of percentages of subject matter in discussions of the ink and stylus texts. Two ink texts and two stylus texts discussed by Expert A and Expert B were used to generate the data

4.3.2 Additional analysis

The Content Analysis of the transcriptions also stressed the difficulty the experts had in reading the stylus tablets, as they concentrated on the lower reading levels: the identification of features, letter forms, and combining them to make words. The higher level analysis (grammar and meaning) they could carry out themselves, but the fact that they could not ascertain the basic features of the stylus texts¹⁴ hindered this process. Analysis of word frequency in the transcriptions confirms this: when discussing the ink texts, the experts most frequently use declarative words about the documents, such as HORIZONTAL, BOLD, FORMAT, DISCOLORATION, HYPOTHESIS, and REASON. When discussing the stylus texts, they use emotive words such as AFRAID, ASSUME, CONFUSING, CONVINCING, DECIDING, SURPRISED, and TRIED. The difficulties in reading the stylus texts can also be seen when comparing the types of subject experts refer to in the discussions, as shown in Fig. 4. They refer more to the feature and characters of the stylus texts than the ink texts, suggesting that they have difficulty in this area, and as a result are unable to draw firmer conclusions regarding the identification of words, and so cannot discuss at any meaningful length, the meaning of the document. This would suggest that the most helpful contribution a computer system could make to the process of reading a stylus text would be to aid in the identification and sorting of data on the feature, character, and word levels (and this, in fact, is the approach adopted in Section 5).

4.4 The model: how experts read ancient texts

Constructing a model (Fig. 5) of a process has two benefits. First, if a knowledge engineer can describe a process in this manner, and the experts agree, it suggests that the process has been understood.

¹⁴ Uncertainty is very seldom fully resolved when reading ancient texts. As shown in the example transcript of stylus tablet 836, there are often characters that are marked by the papyrologists as being uncertain: textual analysis of the corpus of 26,000 characters from the Vindolanda ink texts indicates that 9.9% of the characters in this corpus are highlighted as being uncertain (Terras, 2002). Bowman and Tomlin (2004) provide examples of readings which have changed dramatically between different versions of published texts.

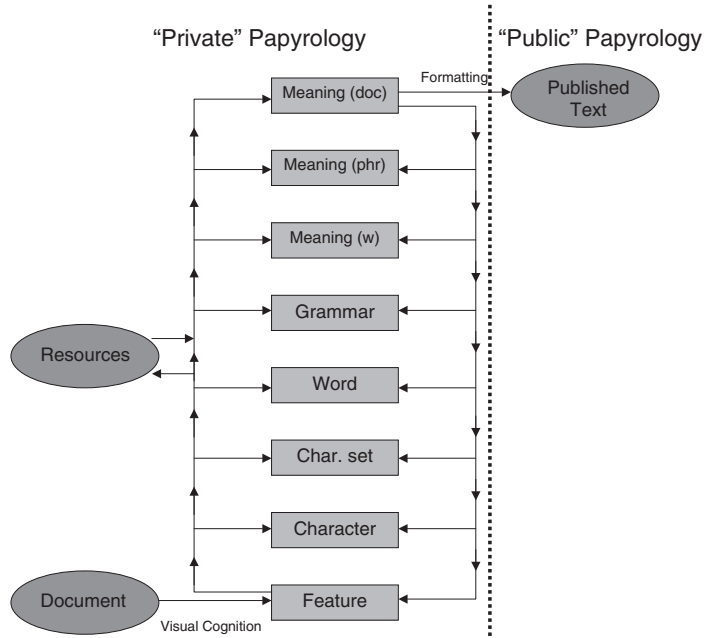


Fig. 5 The proposed model of how experts read an ancient text. Public papyrology (Youtie 1963) refers to the published reading of texts in a common format, after the reading has taken place. Private papyrology is the implicit process the experts undertake when reading a text

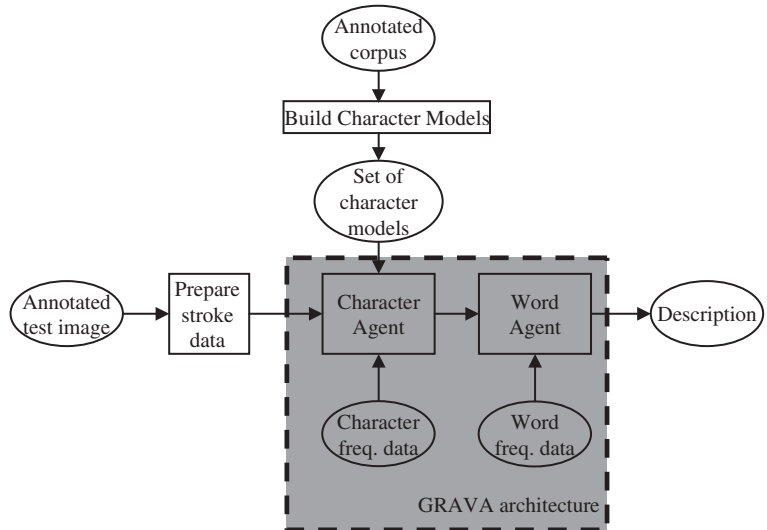


Fig. 6 Basic schematic of system. Robertson’s GRAVA architecture is highlighted to indicate the processes which are carried out as part of the final run of the system

Secondly, such a model can be used as the basis to build an agent-based computer system that replicates the process. The Knowledge Elicitation exercises demonstrated that the experts use a recursive reading mechanism that oscillates between different levels, or modules,

of reading, and the process was rationalized into defined units, to develop a connectionist model of how papyrologists approach and start to understand ancient texts.

An expert reads an ancient document by identifying visual features, and then incrementally building up knowledge about the document's characters, combinations of characters, words, grammar, phrases, and meaning, continually proposing hypotheses, and checking those against other information, until s/he finds that this process is exhausted. At this point a representation of the text is prepared in the standard publication format. At each level, external resources may be consulted, or be unconsciously compared with the characteristics of the document. Although a simple representation, the model shows the overall scope of the process of reading an ancient text: further, more detailed procedural analysis regarding every level (or 'agent') in this system is presented in Terras (2002).

5 Developing an AI-based System

AI has repeatedly shown that interpreting our world requires bringing to bear a great deal of world knowledge. The process of reading an ancient document, especially documents that are in a very bad state of disrepair such as the Vindolanda stylus tablets, is an especially good example of the need to mobilize a great deal of *a priori* knowledge. It is clear that the remnants of writing on the tablets themselves contain insufficient information to recover the original written text, but by applying sufficient linguistic knowledge of character shapes, word and letter frequency, and grammatical information, the tablets can sometimes be read. Implementing a system that can automate the process requires an architecture that can fuse the many different kinds of knowledge in order to arrive at an estimate of the most probable interpretation. Numerous architectures have been proposed to solve interpretation problems: the HEARSAY system (Erman *et al.*, 1980) was developed for speech recognition and more recently Hidden Markov Models (HMM) have been used in natural language processing (NLP) (Charniak, 1993), but these approaches have drawbacks. Blackboard systems can be hard to control, and HMMs can be too restrictive to incorporate complex and diverse kinds of knowledge (Robertson, 1999, 2001; Robertson and Laddaga, 2003).

To construct a computational model of the papyrology process, we adopted, and adapted, the GRAVA (Grounded Reflective Adaptive Vision Architecture) architecture, developed by Dr Paul Robertson,¹⁵ which was built to provide a flexible backbone upon which image interpretation problems can be solved. The GRAVA system was originally developed to analyse aerial satellite images (Robertson, 1999, 2001). The GRAVA architecture uses minimum description length combined with Monte Carlo Select algorithms to find a fair way of comparing data across semantic levels, and is implemented in YOLAMBDA, a dialect of LISP. To test his system, Robertson

15 Formerly of the Robots Group, Department of Engineering Science, University of Oxford. Research Scientist at Massachusetts Institute of Technology's Computing Science and Artificial Intelligence Laboratory (CSAIL)

constructed a small program that could effectively ‘read’ a nursery rhyme, utilizing data regarding letter form, letter frequency, and word frequency, to propagate the most likely interpretation of the image of the text. This system was considerably adapted to work with the Vindolanda data, and the architecture and mathematical underpinnings of the adapted system developed for this project are described in detail in Terras and Robertson (2005) and Terras and Robertson (2004). It is enough, here, to show a model of the system architecture (Figure 6), which demonstrates that it replicates the lower modules of the papyrology process, using AI techniques to sort through the data to propagate the most likely interpretation of the image.

All agents work with data generated from the Vindolanda corpus, as discussed in Section 2: the Character Agent uses the set of character models plus data regarding the frequency of letters in the Vindolanda Corpus, whilst the Word Agent utilizes a list of words generated from the documents read at Vindolanda so far.

To input an image of an unknown document, this image has to be annotated (either manually or automatically, using the image processing algorithms) to identify the key features, and the stroke data prepared for comparison. The stroke data is then passed on to the Character Agent, which compares this data to the character models (generated previously from the Annotated Corpus), and calculates a description length based on which character it is most likely to fit, and the frequency of characters which appear in the corpus (‘i’ is much more likely to appear than ‘q’, for example). One of these characters is selected by a random process, and is passed, with its description length, to the Word Agent. As characters are passed to the Word Agent, they form a string of characters, these resulting ‘words’ are then compared with known words in the corpus, and a description length is calculated for how well they fit known words from Vindolanda. The description length from the Character Agent and the Word Agent is then added, giving a global description length for that run of the system. Subsequent runs of the system generate different global description lengths, as the random process in selecting which letters are passed to the Word Agent assures that different combinations of characters that may fit are put forward as possible interpretations of the problem. When the description length of these separate runs of the system are compared, the Minimum Description Length gives the overall most likely interpretation of the test image. The system has effectively sorted through the different permutations of the data, and provided the experts with the most likely solution to the problem.

5.1 The system in action

When implemented in practice, the system outputs the most likely interpretation of the texts, as can be seen in Figs 7 and 8. An annotated image is entered into the system, and after eight iterations, the system outputs the correct interpretation of the text, keeping note of the alternative possible hypothesis it has also thrown up.

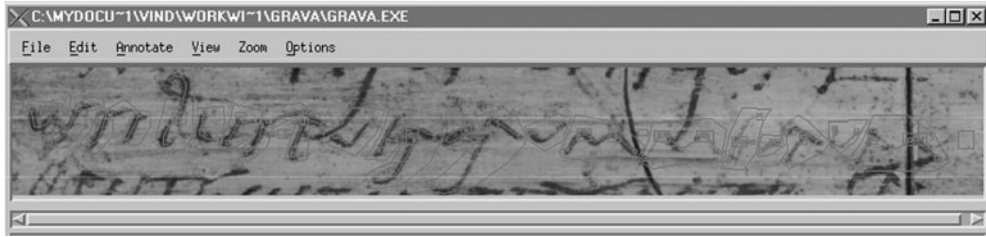


Fig. 7 A section of ink tablet 255, which reads 'ussibus puerorum meorum',¹⁶ used as input

```

Grounded Reflective Agent Vision Architecture (GRAVA) Version 2.0
Yolambda listener pushed. Type: exit to return to GRAVA

=>... load the system and the data ...

=> (runCycles 25)
iteration 0 DL=440.220794 interpretation = ( ... ((2482 252) (2517 250)) ... )
iteration 1 DL=64.085075 interpretation = ( u r s i b u s p u e r o r u m m n o a u m )
iteration 2 DL=49.374412 interpretation = ( u s s i b u s p u e r o r u m m n o r u m )
iteration 3 DL=48.831413 interpretation = ( u s s i b u s p u e r o r u m m n o a u n )
iteration 5 DL=47.816696 interpretation = ( u s s i b u s p u e r o r u m m e o a u m )
iteration 8 DL=36.863136 interpretation = ( u s s i b u s p u e r o r u m m e o r u m )
iteration 25

=> : exit

```

Fig. 8 Output from first successful run on the section of 255, indicating that the system resolves upon the correct interpretation of the image

More remains to be done to make this a useful tool for the papyrologists. The word lists contained within the system need to be expanded, to allow for fragmentary sections of words that may not appear in the existing corpus. The process of annotating a document and feeding it into the system is time consuming, and requires an understanding of the annotation tools themselves. The system will require much development if it is ever to become a stand-alone desktop application, as it currently depends on the knowledge engineer and engineering scientist being present to run it. Nevertheless, it is a working tool that can take in images of texts and output plausible interpretations of those texts, which is speeding up the process of reading the Vindolanda stylus texts. Other types of texts could also be analysed using this system, provided the statistical data regarding letter forms, word frequency, and letter frequency could be obtained.

16 Meaning 'for the use of my boys', referring to some cloaks and tunics for Clodius Super's *pueri*.

6 Conclusion

It has been demonstrated that utilising Knowledge Elicitation techniques can provide the means to understand complex processes in the humanities. This study has made explicit the process involved in reading an ancient text, which has never been properly considered. The cyclic reading process was revealed; the process is not linear, building up one character at a time, but depends on the propagation of hypotheses, and the testing of these regarding all available information concerning a text. Reading a document is a process of resolution of ambiguity, and depends on the interaction of all the different facets of knowledge available to the expert, and this is described in the connectionist model proposed for how experts read ancient texts. Implementing a computer system that works in a similar fashion can then be used as a tool to speed up the reading of damaged and deteriorated texts.

Adopting the GRAVA system provided an architecture that could represent the hierarchical nature of the papyrologist model, and was a successful basis to construct a working computer system that takes in images of ancient documents and generates plausible interpretations of the text of the documents. There is no reason why this system cannot be expanded to encompass other types of documents, if the necessary statistical information was available.

Undertaking a study of humanities-based expert behaviour using Knowledge Elicitation techniques has demonstrated the complexity of the processes undertaken by these experts, and specifically where computer tools can aid them in their task. Although a time-consuming exercise, it demonstrates the level of understanding necessary to construct useful cognitive tools for humanities experts. Only in fully understanding tasks undertaken can computer systems be developed to aid experts in these tasks. Developing this tool has increased the speed with which experts can read ancient documents: increasing the primary data available to other scholars in many different fields who depend on ancient documents for source material.

Acknowledgements

Thanks must go to the supervisors of my doctoral thesis, from which this paper stems: Professor Alan Bowman (Centre for the Study of Ancient Documents, University of Oxford) and Professor Mike Brady (Department of Engineering Science, University of Oxford). The GRAVA system was adopted and adapted with the consent and help of Dr Paul Robertson, formerly of the Robots Research Group, Department of Engineering Science, University of Oxford, and this research could not have come to fruition without his help and generosity. Thanks must also go to the anonymous papyrologists guinea-pigs who gave considerable time and effort to allow this study to take place.

References

- Aalto, P.** (1945). Notes on methods of decipherment of unknown writings and languages. *Studia Orientalia, Edidat Societas Orientalis Fennica* XI.4.
- Aiken, L. R.** (1971). *Psychological Testing and Assessment*. London: Allyn and Bacon.
- Bearman, B. H. and Spiro, S.** (1996). Archaeological applications of advanced imaging techniques. *Biblical Archaeologist*, 59(1): 56–66 .
- Bowman, A. K.** (1997). *The Vindolanda Writing Tablets*. Rome: XI Congresso Internazionale di Epigrafia Greca e Latina.
- Bowman, A. K., Brady, J. M., and Tomlin, R. S. O.** (1997). Imaging incised documents. *Literary and Linguistic Computing*, 12(3): 169–76.
- Bowman, A. K. and Thomas, J. D.** (1983). *Vindolanda: The Latin Writing Tablets*. London: Society for Promotion of Roman Studies.
- Bowman, A. K. and Thomas, J. D.** (1994). *The Vindolanda Writing-Tablets (Tabulae Vindolandeses II)*. London: British Museum Press. Forthcoming.
- Bowman, A. K. and Thomas, J. D.** (2003). *The Vindolanda Writing-Tablets (Tabulae Vindolandeses III)*. London: British Museum Press.
- Bowman, A. K. and Tomlin, R. S. O.** (2005). Wooden stylus tablets from Roman Britain. In Bowman, A.K. and Brady, M. (eds), *Images and Artefacts of the Ancient World*. Oxford: Oxford University Press.
- Brady, M., Pan, X., Terras, M., and Schenk, V.** (2004). Shadow stereo, image filtering and constraint propagation. *Images and Artefacts of the Ancient World*. London: British Academy.
- Charniak, E.** (1993). *Statistical Language Learning*. Cambridge, MA: MIT Press.
- Danks, J. H., Shreve, G. M., Fountain, S. B., and McBeath, M. K.** (1997). *Cognitive Processes in Translation and Interpreting*. Thousand Oaks, CA: Sage Publications.
- Diaper, D.** (1989). *Knowledge Elicitation: Principles, Techniques and Applications*. Chichester: Ellis Horwood.
- Duncker, K.** (1926). A qualitative (experimental and theoretical) study of productive thinking (solving of comprehensible problems). *Pedagogical Seminar*, 33: 642–708.
- Ericsson, K. A. and Simon, H. A.** (1993). *Protocol Analysis, Verbal Reports as Data*. Cambridge, MA: MIT Press.
- Erman, L. D., Hayes-Roth, F., Lesser, V. R., and Reddy, D. R.** (1980). The HEARSAY-II speech understanding system: Integrating knowledge to resolve uncertainty. *Computing Surveys*, 12(2): 213–253.
- Fink, R. O.** (1971) *Roman Military Records on Papyrus*. Case Western Reserve University.
- Holsti, O. R.** (1969). *Content Analysis for the Social Sciences and Humanities*. Reading, MA: Addison-Wesley.
- Kiraly, D. C.** (1997). Think aloud protocols and the construction of a professional translator self-concept. In Danks, J. H., Shreve, G. M., Fountain, S. B., and McBeath, M. K. (eds), *Cognitive Processes in Translation and Interpreting*. Thousand Oaks, CA: Sage Publications.

- Krippendorff, K.** (1980). *Content Analysis: An Introduction to its Methodology*. London: Sage Publications.
- McGraw, K. L. and Harbisson-Briggs, K.** (1989). *Knowledge Acquisition: Principles and Guidelines*. London: Prentice-Hall International Editions.
- Molton, N., Pan, X., Brady, M., Bowman, A. K., Crowther, C., and Tomlin, R.** (2003). Visual enhancement of incised text. *Pattern Recognition*, **36**: 1031–43.
- Pan, X., Brady, M., Bowman, A. K., Crowther, C., and Tomlin, R.** (2004). Enhancement and feature extraction for images of incised and ink texts. *Image and Vision Computing*, **22**(6): 443–51.
- Renner, T.** (1992). *The Wooden Finds from Campania and Dacia as Parallels to Archives of Documentary Papyri from Roman Egypt*. Copenhagen Congress paper.
- Robertson, P.** (1999). *A Corpus Based Approach to the Interpretation of Aerial Images*. IEE IPA99, Manchester.
- Robertson, P.** (2001). *A Self Adaptive Architecture for Image Understanding*. DPhil, Engineering Science, University of Oxford.
- Robertson, P. and Laddaga, R.** (2003). An agent architecture for information fusion and its application to robust face identification. *Proceedings of the 21st International Conference on Applied Informatics*, Innsbruck, Austria.
- Schenk, V. U. B.** (2001). *Visual Identification of Fine Surface Incisions*. PhD, Department of Engineering Science, Oxford University.
- Shadbolt, N. and Burton, M. A.** (1990). Knowledge Elicitation Techniques—some experimental results. In McGraw, K. L. and Westphal, C. R. (eds), *Readings in Knowledge Acquisition, Current Practices and Trends*. London: Ellis Horwood Ltd, pp. 17–23.
- Smagorinsky, P.** (1989). The reliability and validity of protocol analysis. *Written Communication*, **6**(4): 463–77.
- Stemler, S.** (2001). An overview of content analysis. *Practical Assessment, Research and Evaluation*, **7**(17) available online at <http://pareonline.net/getvn.asp?v=7&n=17> (accessed 20 December 2004).
- Terras, M.** (2002). *Image to Interpretation: Towards an Intelligent System to Aid Historians in the Reading of the Vindolanda Texts*. DPhil, Engineering Science, University of Oxford.
- Terras, M. and Robertson, P.** (2004). Downs and acrosses, textual markup on a stroke based level. *Literary and Linguistic Computing*, **19**(3): 397–414.
- Terras, M. and Robertson, P.** (2005). *Image and Interpretation: Using Artificial Intelligence to Read Ancient Roman Texts*. *Human IT*, **7**(3), available online at <http://www.hb.se/bhs/ith/humanit.htm>
- Turner, E. G.** (1968). *Greek Papyri, An Introduction*. Oxford: Clarendon Press.
- Waterman, D. A.** (1986). *A Guide to Expert Systems*. Reading, MA: Addison-Wesley.
- Youtie, H. C.** (1963). The papyrologist: artificer of fact. *Greek Roman and Byzantine Studies*, **4**: 19–32.
- Youtie, H. C.** (1966). Text and context in transcribing papyri. *Greek Roman and Byzantine Studies*, **7**: 251–8.