The Use of Word Level Cues for Script Recognition

Gareth John Bellaby

A thesis submitted in partial fulfilment of the requirements of the Nottingham Trent University for the degree of Doctor of Philosophy

2000

The Use of Word Level Cues for Script Recognition

Gareth John Bellaby

Abstract

Script recognition systems require the use of context to disambiguate input efficiently. One important consequence of the variability of cursive handwriting is ambiguity. Cues at the word level are shown to influence the performance of human readers. A way to build the factors which produce this superior context effect into a machine system is described. This method of using word level cues is called the word level method.

Different, but complementary, sources of information have been integrated to create a robust and accurate machine system. Specifically, a conventional character based pattern recognizer and a word level method have been successfully amalgamated. Work on the integration of information taken from the meta-word level (semantic and syntactic) is also described.

Integration has provided the opportunity to develop interactive processes within the machine system. . This system successfully integrates top-down, and bottom-up, processes. The results of integration for one test sample show an increase in the proportion of target words top ranked from 61% to 70%, in the top 10 from 71% to 83%, and in the top 100 from 72% to 90%. These results demonstrate the efficacy of the word level method and result in a system which has greater scope for improvement when using higher level context.

The work described in this thesis is the author's own, unless otherwise stated, and it is, as far as he is aware, original.

For my wife
Deborah Bellaby
and for my father
Arthur Bellaby

with love and gratitude

Table of Contents

List of Figures

List of Tables

Glossary of Terms

26 word data set. Data set constructed from 26 words, written by one subject, with each letter of the alphabet contained in at least one of the words. See section 3.2.2 and Appendix A.

200 word data set. Data set constructed from writing samples from 18 subjects, each required to write the same 200 words using lower case cursive handwriting (a total of 3,600 samples). See section 4.3.6.2, Appendix D and Appendix E.

Candidate list method. A method of deriving word level cues by using the candidate list generated by the pattern recognizer. See section 4.3 and section 5.2

Catastrophic failures. The pattern recognizer can, on occasion, completely fail to generate any candidates at all. These occasions are called catastrophic failures. There were a total of 512 cases of catastrophic failure in the 200 word data set. See section 4.3.6.2.

Common lexicon. A 15,000 word lexicon, which was created by taking the 15,000 more frequent words from the Lancaster-Oslo-Bergen (LOB) Corpus of British English. See section 3.2.2.

Direct cue extraction. A method of deriving word level cues by using pattern recognition methods. See section 5.3.

IWLM. Abbreviation for the word level method without word frequency information.

Letter verification recognizer. A method of letter verification applied to the word shape recognizer and to the word level method. See section 6.5.

Partial data set. Data set derived from the 200 word data set. Catastrophic failures and uninformative word lists were excluded, and only those lists in which the target word did not appear, or was placed below rank 3, were included. The number of target word in the partial data set was 607. See section 5.2.1.

Pattern recognizer. A recognizer which combines word segmentation and letter recognition. The lexicon is used to prune superfluous work.

PR. Abbreviation for the pattern recognizer.

PR+WLM. Abbreviation for the word level method merged with the pattern recognizer.

PR+WLM+LV. Abbreviation for letter verification applied to the word level method merged with the pattern recognizer.

Uninformative word lists. The pattern recognizer produced a proportion of output which was judged wholly uninformative. These cases, 6 in total, were therefore removed from the 200 word data set. See section 5.2.1.

WLM. Abbreviation for the word level method after the application of word frequency information.

WLM+LV. Abbreviation for letter verification applied to the word level method.

Word level method. A method of applying word level cues. The subject of this thesis but see, in particular, section 4.4 and Chapter 6.

Word shape recognizer. An approach to recognition where an attempt is made to recognize the entire target word, without first segmenting it.

WSR. Abbreviation for the word shape recognizer.

WSR+PR. Abbreviation for the word shape recognizer combined with the pattern recognizer.

SYN. Abbreviation for the merged lists sorted by syntactic class.

Vimes sat back, enjoying a moment's peace.

Something inside his coat went: 'Bing bing bingley bing!'

He sighed, pulled out a leather-bound package about the size of a small book, and opened it.

A friendly yet slightly worried face peered up at him from its cage...

Vimes sighed inwardly. He had a notebook. He took notes in it. It was always useful. And then Sybil, gods bless her, had brought him this fifteen-function imp which did so many other things...

'I think I'll write it in my notebook, if you don't mind,' said Vimes.

'Oh, well, if you prefer, I can recognize handwriting', said the imp proudly. 'I'm quite advanced.'

Vimes pulled out his notebook and held it up. 'Like this?' he said.

The imp squinted for a moment. 'Yep,' it said. 'That's handwriting, sure enough. Curly bits, spiky bits, all joined together. Yep. Handwriting. I'd recognize it anywhere.'

Terry Pratchet, *Feet of Clay*.

Chapter 1: Introduction

1.1 Statement of the Problem

Cursive handwriting is characterized by character strings with ambiguous boundaries and considerable variations in letter form. In consequence, the physical pattern can be interpreted in a number of different ways. This means that any pattern recognizer which is working in anything but the simplest of environments will be unable to completely disambiguate all of its input. Simply put, the problem is that it is not possible to unambiguously recognize cursive handwriting from the pattern alone. A proportion of the output of a pattern recognizer will therefore be ambiguous (target identified, but not selected as the top ranked choice) or incorrect.

A solution to the problem of ambiguity is the exploitation of information found at the word level, and from other contextual sources, integrated together in an efficient manner, in order to improve the accuracy of recognition. The approach proposed in this work can improve upon a set of ambiguous or incorrect results which have been generated by a pattern recognizer as candidates for recognition. The method uses word level cues in order to construct a new list of words. Unlike a conventional character-based pattern recognizer, the word level method does not work on the basis of the physical evidence. Instead it uses lexical information and abstract word level cues. Probable values for certain word level cues are calculated and these values are used to search a lexicon in order to construct a new list of recognition candidates. Viable candidates can be derived in this way, even when the conventional pattern recognizer did not identify the target word. No attempt is made to edit the candidates generated by the pattern recognizer; rather, new, probable alternatives are generated. The new candidate list is then merged with the output of the pattern recognizer. The result of this is that system accuracy and robustness are both improved. The word level method is not a replacement for pattern recognition but rather acts as a different knowledge source which can be combined with pattern recognition in order to improve machine performance. Although the conventional pattern recognizer and the word level method may draw upon similar sources of information, the information is being used in completely different ways.

One way of depicting the human reading process is to make a distinction between the letter level, the word level and the meta-word level and to consider the different sources of information which can be utilized at each of these levels. This can be seen in Figure 1-1.

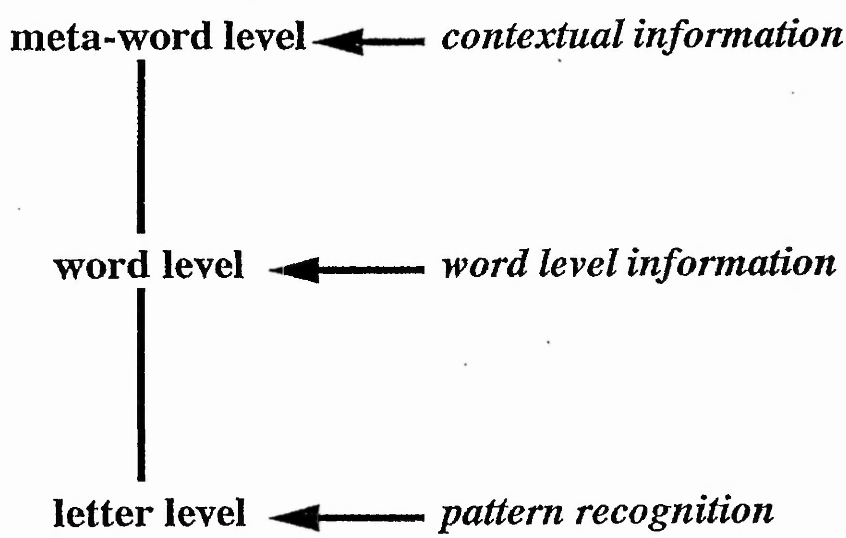


Figure 1-1: One view of the reading process

The letter level is where the basic recognition of individual characters occurs. The method used to identify individual letters is that of pattern recognition. Physical cues are used by the reader in order to try to identify the letters which he or she is attempting to read.

The word level, in contrast, is where the reader attempts to recognize whole words, rather than their individual constituents. Pattern recognition can be done at the word level, but it is the stage where human readers also introduce additional information. Human readers use word level information in order to aid their recognition of individual words. This information may involve the use of information which has a physical component, such as likely letter candidates, but it also involves the use of non-physical cues, for instance the use of lexical information to combine these letters in order to form words. The word level, and the use of word level information in order to improve the machine recognition of cursive script, is the subject of this thesis.

The meta-word level is where the reader uses information drawn from a word's surrounding context as an aid to recognition. A human reader can use cues from a word's neighbours, the clause or sentence in which a word occurs, and the text in which a word is embedded, to help his or her recognition of the word. Meta-word information includes such sources of information as syntax and semantics.

Although the reading process has been depicted as a hierarchical series of layers, information probably travels down through these layers as well as up. For instance, the identification of a word may involve both physical information about its constituent characters drawn from the letter level and contextual information drawn from the meta-word level about likely letter and word candidates.

There are limits to the recognition of isolated letters, i.e. letters without any supporting contextual information. Cues at the word level influence the performance of human readers. These cues can be applied and integrated into the recognition process in order to improve machine performance.

Words in isolation, i.e. without any surrounding semantic or syntactic information, are not always recognized perfectly, even by competent adult readers. Meta-word contextual information must be used to disambiguate input in such cases, although the addition of such information will not guarantee completely accurate recognition. One important topic of this thesis is the way in which pattern level, word level and meta-word level sources of information can be integrated to produce an efficient and robust recognition system.

1.2 Motivation for Text Recognition

There is great interest in the development of a machine interface which can use a natural means of communication. One standard mode of communication is writing. Increasing attention has been shown to handwriting recognition systems since there appears a greater likelihood of being able to construct a robust system which can deal with direct written input than there is for spoken language recognition, the closest rival to handwriting recognition.

It is true that it is possible to type more quickly than it is to write but writing is learnt at an early stage in life, whilst typing is learnt, if it is learnt at all, at a later stage. Writing is an extremely widespread mode of communication: more people are capable of writing than are capable of using a keyboard. Script recognition systems are a better alternative than speech recognition systems in situations where there is likely to be a lot of noise, thus making speech recognition more difficult, or where a greater degree of privacy is required than can be offered by speech recognition. An area where handwritten input is likely to prove useful is in cases where a keyboard is unsuitable, such as is the case with small, highly portable, computer systems. There already exist commercial pocket sized computers. Such computers are too small for adequate keyboards to be used easily. In the case of such computers, handwritten input can provide an ideal interface. In general, script recognition systems are ideal for non-standard input that requires characters or symbols not found on a standard QWERTY style keyboard. Finally, handwriting recognition systems provide a direct relation between pointer and writing implement: both are the same instrument. This can be contrasted with the by now standard mouse and keyboard set-up in which writing and pointing are two separate activities requiring different equipment. Handwriting recognition systems provide a more natural and intuitive interface with the computer.

1.3 Characteristics of Cursive Handwriting

1.3.1 Ambiguity

The recognition of characters is not a simple task. The English alphabet is composed of 26 letters, each of which has both an upper and a lower case form. There are also 10 numerals and several punctuation marks. Some of the punctuation marks are well fixed in the language, play an important syntactic role and are widely used (e.g. full stop, comma, apostrophe, question mark, exclamation mark). There are roughly 80 different basic patterns which a reader has to recognize. However, there are also many other characters or symbols in the language (e.g. mathematical symbols). These characters often have a very specific use or meaning. New instances of such characters are not unusual since they are often coined as shorthand.

However, it is not the number of patterns to be identified which is the central problem in the recognition of cursive handwriting, but rather their ambiguous nature. A central argument of this thesis is that natural cursive script is inherently ambiguous and that a machine system cannot be expected to recognize individual letters beyond a limited degree of accuracy by pattern recognition alone. Ambiguity may be present in handprinted characters or in sufficiently degraded typeface characters, but its effect is pervasive, and extreme, in cursive script.

The ambiguity of cursive handwriting is, firstly, the result of uncertainty. One pattern can signify more than one character. It is common for some letters to be written in a very similar fashion. For example, the next to last letter of the word shown in Figure 1-2 may be an 'r' but it could also be a 'v'.

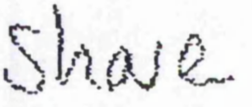


Figure 1-2: A handwriting example

Secondly, ambiguity is caused by variability. Characters are just a set of arbitrary strokes whose variability between people, and even within the handwriting of one person, can be substantial. In the course of any composition, a writer will vary the way in which he or she writes a letter. The way in which a character is written can be affected by its surrounding characters, e.g. the letter 'e' in "cent" is often written differently to the 'e' in "rent" [Eldridge, et. al., 1984]. (See Figure 1-3.) It is not unusual to see a writer using two separate letter forms for the same character even in a single word. The degree of variability apparent in written letter forms is vastly increased when the handwriting of different individuals is considered. Several examples showing these characteristics of handwriting are presented in Figure 1-4, Figure 1-5 and Figure 1-6. Each person has his or her own style of writing to the extent that it is often possible to recognize the identity of the writer from his or her handwriting.

Figure 1-3: A handwriting example

Figure 1-4: Two examples of the same word by one writer

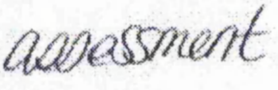


Figure 1-5: Two forms of the same letter in one word

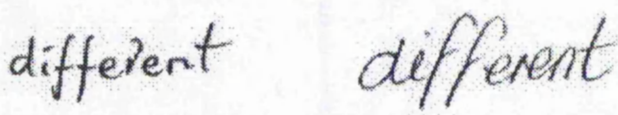


Figure 1-6: The same word written by two different writers

Letters run together in cursive script. Separating the letters is called "character segmentation". Cursive script recognition is particularly difficult because several characters can be written with a single stroke, and because there is great variability between writers in the spacing of individual letters. Characters in cursive script can touch and they are often joined together by connecting lines. These extra connecting lines are called "ligatures". Character recognition is dependent upon where the character boundary is placed. A large number of alternative segmentations may be possible for a word. The task of identifying characters in cursive script is therefore combinatorially complex.

It is the case that the same written pattern can represent more than one word. For example, the word given in Figure 1-7 can be read as "dog", but it could also signify the word "clog" or even "cloy". In other words, this pattern of lines on the page can legitimately signify at least three different words.

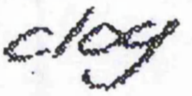


Figure 1-7: A handwriting example

1.3.2 Redundancy

Redundancy is a characteristic of writing and speech in general. The high degree of redundancy in writing is intended to allow meaning to be drawn from a few pieces of the original communication. Put simply, the more redundancy a signal contains, the less likely it is that its meaning will be lost. Redundancy decreases the importance of any one unit of a signal, to the signal as a whole.

The English character set and writing conventions compensate for ambiguity and variability, at least in part, by the duplication of information and by the use of supporting data. For example, the letter 'i' can be distinguished from the letter 'l' by the length of its vertical stroke, but discrimination is supported by a dot being placed near the 'i'. The break between one sentence and another is signified by three separate items of information: a full stop, a space and the use of a capital letter at the beginning of the new sentence.

A significant number of words in ordinary writing are not needed for the text to be understood. For instance:

> Omit much words make text shorter. You not have trouble understanding.

Likewise writing can be degraded using other means and yet its meaning will still be apparent:

> Thxs wx cax rexlaxe exerx thxrd xetxer xitx an x, axd yxu sxilx maxagx prxttx wexl. Thng ge a ltte tuger f w sipl deet th lete.

Or:

> Thus we can replace every third letter with an x, and you still manage pretty well. Things get a little tougher if we simply delete the letter. [examples from Lindsay and Norman, 1977]

If written language was more efficient then readers would need to attend carefully to every word and letter presented to them, since one mistake might distort the meaning of what they are trying to read.

Writing conventions have developed within two important constraining factors: the ease with which words can be written, and the ease with which they can be read. The requirements of both readers and writers mean that handwriting is not highly efficient. Hence, the introduction of redundancy is the only way in which the meaning of the text can be sustained.

Redundancy allows language to be a flexible medium for communication. In consequence good clear writing, such as a piece of printed text using a legible font, can be read rapidly, whilst less legible writing, such as poor handwriting or a briefly scribbled note containing abbreviations, may be read at a slower pace but still be understandable. Redundancy plays an important part in communication. Redundancy allows a reader to attend selectively to different possible features of a text, e.g. to look for key words and phrases. It also permits a reader to anticipate what will come next.

One further consequence of the existence of redundancy in language is that it also allows a human writer or speaker in certain circumstances to introduce more than one meaning into a sentence, for instance to produce a pun or to make an ironic statement.

English also displays a high degree of orthographic redundancy. There is a sequential redundancy of letter strings in English orthography. If there were no statistical redundancies between letter positions this would mean that the probability that a character will occur is not dependent upon the preceding characters. However, from a given character, or sequence of characters, it is possible to predict what the next character is likely to be.

For example, there are virtually no instances in English where the letter 'q' is not followed by the letter 'u' (one of the few exceptions is "qwerty"). If a reader comes across the letter 'q' then he or she can be very confident that the next character will be 'u'. Therefore, visual identification of the 'u' will provide the reader with little or no extra information. This means that the sequence 'qu' is barely more informative than the letter 'q'.

The fact of this redundancy can be demonstrated statistically. C.E. Shannon estimated the redundancy of English orthography to be 50%. English texts are thus roughly twice as long as they need to be [Shannon, 1948]. This is a consequence of the way in which English is spelt. Shannon suggested that once long range statistical effects (up to 100 letters) were taken into consideration the redundancy of English could reach 75% [Shannon, 1950]. For example, the Oxford English Dictionary contains 290,500 word entries and 616,500 word forms once variant spellings, obsolete forms, combinations and derivatives are included. However, a massive 12,376,630 unique letter strings of 5 letters or less can be generated from an alphabet of 26 letters.

Writing is a human tool. Whilst this tool is used for communication, its function is not the efficient transmission of facts. More efficient forms of written communication can be found, but these are designed for certain specialized functions. The sequential redundancy of general English can, for instance, be contrasted with the more efficient character set and grammar of mathematics or symbolic logic. However, maths and symbolic logic are only small subsets of general writing. Writing has also been fashioned to affect emotions, to engage the writer's and reader's skills in their use of the tool, to allow for new, imaginative and creative uses, and for aesthetic considerations, as well as the conveyance of information.

The level of redundancy in written English has been determined by its users over a great many years. Too great a degree of redundancy would make reading and writing tiresome, whilst the requirements of flexibility (e.g. the need to address audiences of differing reading abilities) and the need to retain meaning even when the visual data is degraded mean that an appropriate level of redundancy has been retained.

The level of redundancy in written English is therefore a compromise between such things as efficiency of communication, the need for such a form of communication to be adaptable to widely different circumstances, the writer's ease of use versus the reader's, the different writing tools available, and the need for the medium to be adaptable to different styles of writing and reading. The relevance of redundancy is that a range of constraints can be utilized by a machine system for the task of script recognition. The fact that language is highly redundant is important for the machine recognition of handwriting since it strongly indicates that the recognition of written text is not solely dependent upon a few limited sources of information. Redundancy means that it is possible for a machine system to successfully utilize sources of information which a human reader may not, or which a human reader only uses to a limited degree. For example, a machine system can use the sequential redundancy of letter strings to calculate the probability of succeeding characters in a way that a human could not.

1.4 Why Examine the Human Reader?

In order to develop a script recognition system it may help to examine the way in which humans read. The reason for examining the way in which humans read is that people are the only efficient readers of unconstrained handwritten text. Since reading is a human activity, it is reasonable to use information about the human reading process and the sources of information which human readers use during these processes. Handwriting has developed with a human audience in mind. Given that a machine system is expected to read normal handwriting, it seems a well-grounded assumption that such a system will have to draw upon similar sources of information to its human counterparts and for its recognition processes to bear some similarity to human reading processes.

However, it is reasonable for a machine system to use analogues which only mimic their human counterparts. It is not necessary for a machine system to read text in the same way that people do. The fact that human readers use a particular approach is not enough to justify its use by the machine, although it is the case that human activities will cast light on promising techniques. Whilst some of the methods finally adopted may borrow from human processing, the objective of the work itself is not to simulate human thought processes. This can be compared to the computer systems developed by psychologists such as McClelland and Rumelhart [McClelland & Rumelhart, 1981]. The aim of developing such systems is to simulate models of human reading in order to test hypotheses about these models and to generate further insights into the human mind. For this present thesis, human readers are being studied in order to gain information about how a machine system can be made more robust and efficient.

It might be assumed that cursive script recognition is one of those areas in which machine performance can outstrip human performance, certainly such a hope has been expressed by some working in the field [e.g. Earnest, 1962]. However, no robust evidence or argument has been presented which suggests that handwriting is still anything but an area in which human readers have a natural advantage. Furthermore, human readers demonstrate a degree of flexibility in their reading which is not yet demanded from machine systems, e.g. the ability to read both handwritten words and printed words.

Human readers have a general reading ability. For the human reader, reading must be a multi-purpose tool. Human readers are required to read about many different subjects under a variety of circumstances. This is not necessarily the case with a machine system. A machine system may be required to recognize only a limited vocabulary which is perhaps presented in a specific and pre-determined format. A machine system may be better than a human reader at the task of recognition when this recognition is to be carried out for a specific limited purpose (e.g. reading a standardized form), or when the subject of the writing is known to lie within a specialized domain.

It is assumed that if a source of information is used by human readers then it can also be used by the machine, i.e. data available to a human reader is also available to a machine system. In one important respect, this is not true. Human readers have access to a vast store of world knowledge which it is not yet possible for a machine system to use. Indeed, it is expected that any foreseeable machine system will be unequal to a human reader in this respect. Putting aside the large amount of memory which would be required to store such information, there is still no agreement about how such knowledge should be represented, organized and processed. A machine system is thus placed at a disadvantage.

Although a source of information is available to a machine system, it still may not be possible for a machine system to use every source of information as efficiently as a human reader. However, a machine system can engage in more detailed analysis and may be more consistent than a human reader. It should be possible to exploit this consistency and attention to detail to compensate for the system's lack of human knowledge. This can be an important factor in the potentially complex task of reading handwritten text. For a machine system to be effective, it needs to be able to access as many sources of information as it possibly can. It is also hoped that a machine system may be able to exploit those sources of information to which it has access more efficiently than a human reader. Such an ability may compensate for the disadvantage which a machine system has of being unable to use the same comprehensive store of world knowledge that a competent human reader can use.

1.5 Performance Issues

Machine recognition systems are generally based mainly or wholly on the pattern recognition level. However, it is clear that it is not possible to get complete disambiguation from pattern recognition alone. A limited machine system (e.g. one intended for use by a single writer) may experience this problem to a much lesser degree, but it will still exist. For instance, two different characters may be written in a similar manner even by one writer, all handwriting displays some variability, and a machine system will face segmentation problems even with clear, neatly formed cursive script [Bellaby & Evett, 1994a].

Very little data on how well the pattern recognition stage of a script recognition system should perform has been accumulated. It is generally accepted that handwriting is ambiguous and that contextual information must be taken into account if the performance of machine systems is to approach that of human recognition. However, there is less discussion about exactly what sort of performance should be expected from pattern recognition software. It can be argued that there are already a number of pattern recognition systems developed whose character identification performance is close to that of human readers, but it is too easy to pay lip service to the notion that an upper limit exists, whilst in practice ignoring the implications of this. Since any cited figure will of course be dependent upon the handwriting data used, it can be assumed that the suggested limit is not applicable [Bellaby & Evett, 1994a]. One way to establish upper limits on performance is to compare different pattern recognition systems. An alternative method is to examine human recognition. This method is used in Experiment 1 (see section 3.2) Machine systems should not be expected to solve the problem of script recognition on the basis of pattern recognition alone; there is a point beyond which further development at this level is futile, and effort should be directed towards the implementation of other sources of information.

Occasionally in cursive script, letters or sequences of letters, appear as just a trailing line, typically at the end or middle of a word. Given such input a machine system can have little or no chance of identifying the character or characters by pattern recognition alone. It is impossible for a pattern recognizer to produce any suitable candidates given sufficiently degraded input. Human readers will be confused when presented with an 'r' which resembles a 'v'. This confusion is a direct result of their need and ability to understand many different kinds of handwriting. It should therefore be accepted that a machine recognition system will experience similar difficulties to those experienced by human readers.

Our knowledge about how humans read is incomplete. However, it is possible to demonstrate that certain sources of information are important to human readers. Specifically it is feasible to study human readers and the reliance which they place on different sources of information. It is possible to suggest figures for how well human readers use certain sources of information and to indicate the degree of influence which the information has upon the reading system as a whole and upon apparent processes within this overall activity. This information can help to show how good a machine system should be in the same areas, and the sources of information which it can, perhaps should, use.

1.6 Levels of Performance on Cursive Script

The difficulties which are apparent in the reading of cursive handwriting are not just a consequence of general legibility. Ambiguity is a necessary consequence of the real variability and uncertainty of written letter forms. Indeed, typeface was derived from handwritten letter forms with the intention of improving character legibility.

Suen has presented some figures on the accuracy of the human recognition of handwritten letters [Suen, 1983]. Three styles of writing were considered: block printing, manuscript writing and cursive writing. The letters were segmented out of words by hand. Suen found that 1.25% of block prints, 2.39% of manuscript writing and 4.73% of cursive writing were illegible to the subjects. Some of the typical confusions made on cursive handwriting were 'c'->'i', 'r'->'i' and 'v'->'u'. Two major criticisms can be made of this experiment. Firstly, no limit was placed on the amount of time a subject could view the stimuli. This means that the subject's response was not necessarily his or her first, immediate impression of the stimuli. A true comparison with the letter recognition time apparent with normal reading speeds has not therefore been made. Secondly, no allowance was made for the misidentification of single letters as double letters (or vice versa). This means that some of the problems of segmentation were also ignored. The error rate for all three styles of handwriting is almost certainly much greater than the figures given. Chapter 3 presents an experiment which takes the above objections into account. This experiment also presents a direct comparison between human and machine performance on cursive script.

Schomaker reports the results of some small-scale experiments in which human readers were presented with cursive words written with ballpoint on white paper. Recognition rates on cursive words ranged from a top rate of 88% words recognized to a low of 54% recognized. The top rate was for single-word recognition of neat, frequently used words all of which had been written by a single writer. The low rate was for the middle word of three unrelated words each of which had been written by different writers, and the word had been written rapidly [Schomaker, 1994].

1.7 The Importance of Meta-word Context

The fact that context facilitates word recognition is well established. A significant part of the interpretation of visual data by a reader is provided by the reader's knowledge of what the data should be, rather than from information contained in the data itself. This extra information is derived from the context in which the visual data is embedded.

For reading this means that whilst the reader needs to pay attention to the actual patterns of the characters on the page, he or she will also use a great deal of knowledge about what is being read.

The reader will know that it is indeed characters which he or she is seeing rather than a set of meaningless marks, that the text (for instance) is written in English, with all that this implies about orthography and the syntax and lexical construction of the text, and lastly the reader will have some awareness of the meaning of the text.

It is not necessary for a human reader to recognize every individual letter or word in order to read a text successfully. Human readers use contextual information, as well as letter recognition, in their recognition of handwriting. This implies that a successful machine system must take both context as well as pattern recognition into account. The existence of ambiguity in handwriting means that an opportunity exists to exploit contextual information. A system with the ability to use context will be more robust and flexible than one that simply uses pattern recognition since it will be able to use this contextual information as an aid towards recognition and to help resolve points of ambiguity at the pattern level by using contextual information to select between alternative interpretations of a written word.

There are two different kinds of meta-word context. Firstly, there is the effect of global context, i.e. the text in which the word is embedded. The effect of global context is, for example, semantic and pragmatic. Semantic effects must be essentially indifferent to syntax since the meaning of a text can be found independently of syntactic variation. Secondly, there is the effect of local context, i.e. the immediately surrounding words. It must be the case that the effect of local context has an important syntactic component. For instance, appropriate syntactic context results in better human performance even when meaningful semantic information is absent, e.g. 'the colourless green ideas dream furiously'. There is a context effect at the word level called the Word Superiority Effect [Cattell, 1886; Baron & Thurston, 1973]. Human readers recognize letters in words more easily than they do letters in isolation or letters in non-words. This is discussed in Chapter 2.

An example of the extent to which context improves the recognition of individual words and letters is the fact that a professional proof reader employs the technique of reading a text backwards in order to reduce contextual effects and so improve his or her awareness of the actual, rather than perceived, spelling of individual words.

There have been many studies showing that context facilitates word recognition. This has been demonstrated using a variety of reading tasks. For instance, the period required to respond to a word embedded in text is strongly influenced by the context preceding that word, e.g. the lexical decision time for the word 'butter' is faster when it follows the word 'bread' than when it follows a word which is unrelated [Meyer, Schvaneveldt & Ruddy, 1975]. Semantic information [e.g. Fischler & Goodman, 1978] and syntactic information [e.g. Miller & Isard, 1963; Goodman, McClelland & Gibbs, 1981] have both been shown to improve human reading performance.

It is not difficult to show the effect of context, both at the word level and at the meta-word level. The sequence of patterns in Figure 1-8 is unrecognizable as a word. However, in Figure 1-9 the same sequence is presented, this time with a mark placed over the patterns and embedded within the context of a familiar phrase. In this instance, it is not difficult to identify the intended word. The mark placed over the sequence patterns serves as 'noise', and since human beings are accustomed to deciphering such noisy or degraded patterns, the mark actually acts as an aid to recognition.

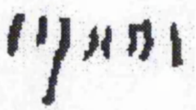


Figure 1-8: An example of a degraded word

Figure 1-9: The same word with noise and in context

Even without the help given by the introduction of overt noise, it should be easier to recognize the pattern given in Figure 1-10 as the word 'sunny' when it is placed in an appropriate context, as in Figure 1-11.

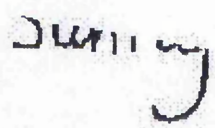


Figure 1-10: A second example of a degraded word

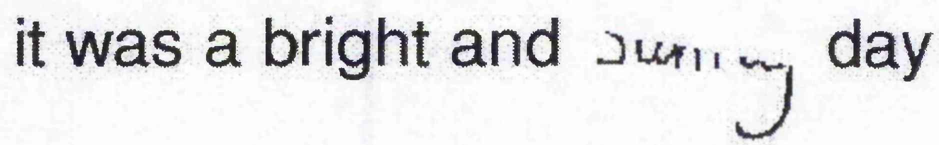


Figure 1-11: The second word in context

Human readers utilize context to the extent that it can in certain cases override the actual patterns written on the page. For example, a reader may be unconscious of spelling mistakes, omissions or repetition. It is not easy, for instance, to spot the mistake in the following example

PARIS IN THE

THE SPRING

The failure to spot this kind of repetition is particularly apparent with function words such as "the". Content words are more likely to be fixated than function words. Contextual evidence suggests that the word "the" should be present in the text. Perceptual and contextual evidence confirm that the word is indeed present. The reader is not conscious that the second word is also present because contextual cues suggest that there is no need for it to be there.

It should be remembered that the goal of the reader is to understand the text, not the individual letters or even words of which it is comprised. Indeed, it is even possible for people to miss out whole sentences and still comprehend the text being read, e.g. speed reading [Just, et. al., 1982].

1.8 System Overview

It is only when contextual factors are added to the pattern recognition process that the machine recognition of unconstrained handwriting will approximate that of human understanding.

Systems incorporating pattern recognition and several higher level knowledge sources have been developed [Boes, et. al., 1989; Wright, 1989; Wells, 1992; Keenan, 1993; Rose, 1994]. Further work on developing new sources of information and extending existing sources is underway [Powalka, 1995].

In brief, the machine system which is the target of the current thesis consists of a conventional pattern recognizer that produces a set of individual characters, together with segmentation information about these characters. These characters are combined to produce letter strings which are filtered according to a lexicon to produce a set of word candidates. Finally, words are combined to form word strings, i.e. to produce a set of sentence candidates. Syntactic and semantic information is used to filter the set of sentence candidates to produce a final choice. See Figure 1-12

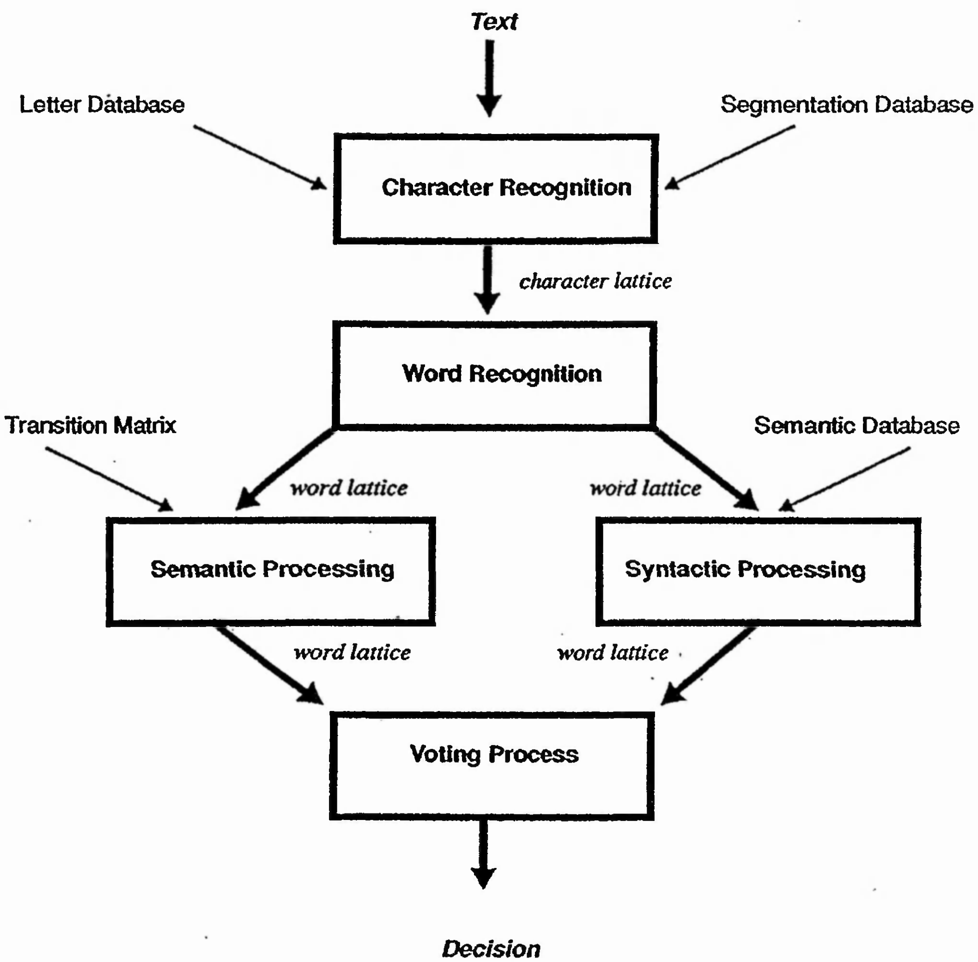


Figure 1-12: The existing machine system

The system is to operate in real time, i.e. as the text is being written [Evett, et. al., 1992]. The objective is a machine system which will overcome the particular problems associated with unconstrained cursive handwriting and a large lexicon.

1.8.1 Data Issues

One purpose of the work being undertaken is the development of a machine recognition system which can deal with unconstrained cursive handwriting, i.e. a writer independent system. It is not intended that writers will have to adjust their style in order to communicate successfully with the machine system, for example, they will not be expected to write extremely clearly or to use block writing.

Current efforts are not directed at the identification of words which have been misspelt, although it is hoped future work will address this problem. No attempt is made to recognize data which contains writer or device errors. It is, in this respect, "clean data". It is necessary to assess the legibility of the data used since it is unreasonable to expect the machine to recognize words which are ill formed. However, it is not easy to classify the legibility of cursive writing. Schomaker, for instance, distinguishes between neat cursive, cursive and fast cursive script [Schomaker, 1994]. Human judges are used to assess the writing before it is presented to any machine system. The data used in the author's work is considered to be "normal" cursive handwriting. That is to say, words written at normal writing speed and not written under abnormal conditions. The writers are specifically requested not to modify their writing style, so the writing collected is not overly neat, nor particularly illegible.

1.8.2 Character and Word Recognition

Some of the particular problems involved in the pattern recognition of cursive script are: the segmentation of a word (i.e. the detection of the start-end points of the characters which make up the word); the fact that the same letter does not always have the same pattern; and the fact that template and stroke pattern will never exactly match, e.g. their lengths will differ.

Characters and words are recognized using a segmentation based character recognizer and a dictionary look-up process [Powalka, et. al., 1993]. This is a strict recognition method. Only those characters which have been explicitly identified by the system are considered to be present in the word. Data are written on electronic paper. The data are recorded as a sequence of x-y co-ordinates. These input data points are then encoded using a vector direction encoding method based on Freeman encoding modified by Powalka [Freeman, 1974; Powalka, 1995]. Adjacent points are converted to vectors. An eight direction vector encoding method is used. The data are segmented into the smallest useful segments (called 'clusters'). The recognizer then attempts to combine clusters into letters. This is done by consulting a segmentation database. The resulting letter pattern is used to derive a number of new patterns which vary in the amount of detail they retain. Letter patterns are then matched against a database of known patterns. In this manner, letter sequences are built up. A lexicon is used to verify the letter sequences and only known letter combinations are processed further [Wells, et. al., 1990]. In practice, an interactive method which combines segmentation, letter recognition and lexical look-up processes is used. Powalka calls this recognition method "multiple interactive segmentation". Figure 1-13 shows an example of the multiple interactive segmentation of the word "me", depicting some of different recognition sequences caused by each possible segmentation of the data.

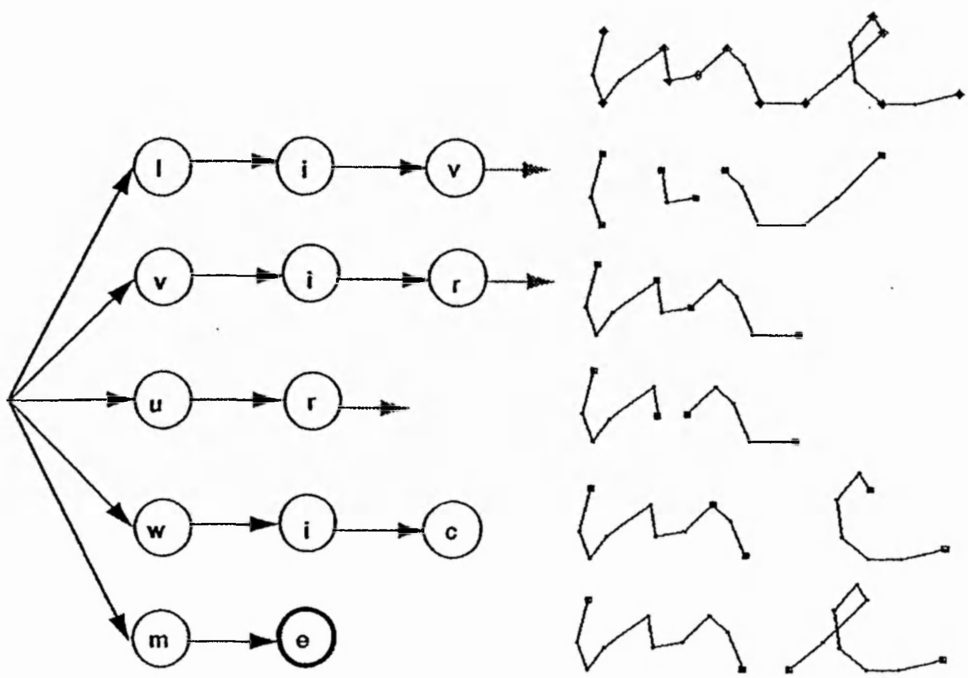


Figure 1-13: Multiple interactive segmentation. Some paths through the lexicon with corresponding segmentations [Taken from Powalka, 1993]

The final outcome of the pattern recognizer is a list of word alternatives. Each word alternative is given a confidence score derived from its physical characteristics. A confidence value is ascribed to each letter that has been identified. The confidence value given to each letter is derived without reference to letter position. This value reflects such characteristics as the size of located letters and their proportions. Dots for the characters 'i' and 'j' are used indirectly by the recognizer. If either 'i' or 'j' has been suggested as a candidate then the presence of a dot is sought in an attempt to confirm the suggestion. If found the confidence value given to the 'i' or 'j' as a candidate is increased. A similar process is used for the presence and position of dashes. The resulting list of words is ranked according to the confidence score which has been given to each of the word candidates. This score is in the range 1 to 100. The maximum possible number of words in the list of alternatives can be varied, up to a limit of 100 alternatives.

A pragmatic approach, rather than a training process, was used to create the initial letter and segmentation databases. Samples of writing were examined. From these observations a set of generic letter forms was created. Permutations of the templates were then generated and verified by examining real data. The pattern recognizer subsequently received some limited training on further data samples.

Alternative pattern recognizers have also been developed by Powalka. The endings of long words are sometimes less legible than the rest of the word. A word ending postulation method which attempts to overcome this difficulty has been developed. Combinatorial complexity is restrained by placing a limit on the number of words which can be postulated. A maximum of five words can be postulated. Word ending postulation is less accurate than strict recognition since more than one ending can be found for each word stem, but the former has a lower overall error rate than the latter. The two methods have been used together to form a joint hybrid recognition method which combines the best characteristics of each method [Powalka, et. al., 1993].

Word ending postulation was a method developed to overcome the residual but significant group of ambiguous words which the conventional pattern recognizer was unable to recognize. Although it met with some success, it was evident that alternative methods would have to be developed.

Recent work has concerned the development of a recognizer which attempts to exploit word shape information, such as zoning information and the location of vertical bars [Powalka, et. al., 1994]. It is argued that the combination of several recognizers each of which is capable of extracting a different characteristic of cursive handwriting is the best route forward [Powalka, 1995].

The work presented in this thesis draws primarily upon data generated using the strict pattern recognizer. A strict match between the input pattern and the output of the recognizer was required as it was felt that postulation and estimation methods might cause invalid assumptions to be made about the input data and recognition method, e.g. the pattern recognizer's accuracy in the task of identifying the last letter of a word. However, pattern recognition methods developed for the word shape recognizer have been used for the purposes of direct cue extraction (see section 5.3).

1.8.3 Word Recognition

It is apparent from examining the way in which humans read that context effects are very strong and that one particular and important point where contextual cues are used is at the word level.

The application of such information to the machine recognition of handwriting forms the basis of this thesis.

1.8.4 Meta-word Information

Two kinds of meta-word contextual information are used: syntax and semantics. In both syntactic and semantic analysis a word lattice is constructed out of multiple word lists, e.g. see Figure 1-14.

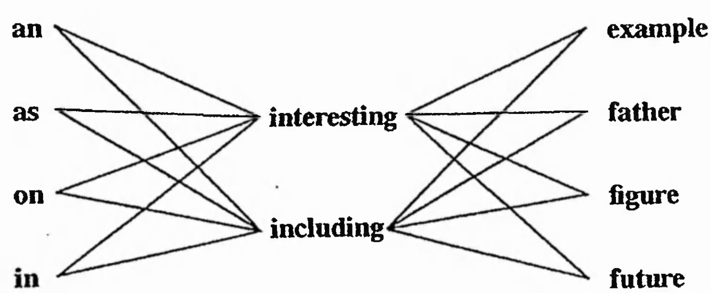


Figure 1-14: A word lattice example

The number of columns to each side of the current candidate list (the window) can be varied. The probability of a tag appearing in the given context is derived from an analysis of the word lattice. The methods which are used for semantic and syntactic analysis are not dependent upon the ordering of the word list.

1.8.4.1 Syntactic Structure

There are two approaches to the syntactic analysis of text by machine systems. Firstly, the use of rule-based parsers based on generative grammars (e.g. generalized phrase-structure grammars, or augmented transition networks) and, secondly, the use of probabilistic analysers. This second approach to syntactic analysis depends on the statistical properties of language structure rather than absolute logical rules. The machine system does not "know" anything about the grammar of English sentences. A "tag" or grammatical code is assigned to each word of an input sentence. The probability of a word appearing in its given context is calculated by reference to a table of empirically-derived relative frequencies of transitions between adjacent tags [e.g. Church, 1988]. The data for the probability estimates are derived from natural language corpora. The grammar tags used in the Nottingham Trent syntactic analyser are a mixture of the Lancaster-Oslo-Bergen (LOB) Corpus of British English tagset [Johansson, 1980] and the Oxford Advanced Learners Dictionary tagset [Keenan, 1993]. The probabilities of successive tags across the lattice are combined to give tag strings ranked in order of their likelihood; associated word strings can then be given in rank order.

It is argued that authentic data provides evidence against any clear-cut distinction between grammatical and 'deviant' sentences in natural language or, to put this another way, "all grammars leak" [Sapir, 1921]. The lack of such a distinction precludes the possibility of successful machine systems which rely on a generative grammar. In contrast, corpus-based linguistics makes no such absolute distinction between well-formed and ill-formed sentences, but merely looks at the relative frequencies of grammatical classes in naturally occurring language [Sampson, 1987]. Probabilistic analysers are also faster than rule-based parsers since the main computational workload for the former occurs offline during the initial analysis of the corpora.

1.8.4.2 Semantic Processing

A text or discourse is not just a collection of unrelated words or a series of sentences, each on some random topic. Rather, the words, sentences and phrases of any coherent text will tend to have some sort of unity of purpose: sentences, phrases and paragraphs function as a unified whole. Semantic processing promotes words which are considered to have a relationship with one another.

An informed estimate of whether there is supporting evidence that a word fits in its given context is used. The current system uses dictionary definitions and collocational information from natural language corpora [Rose, 1994]. Collocational information is derived by calculating how often words are found together in a corpus. A collocational "definition" is then recorded for a word. Words in the lists of alternatives are given a score based on how well they combine with neighbouring words in the input. If the dictionary or collocational definition of one word contains the other word then each of the two words is given a high score. A lower score is ascribed to each of the two words if their definitions contain a number of similar words. If no definitional overlap between a pair of words is detected, then no score is awarded. It is therefore possible that a word receives no score from the semantic analyser. Function words are also ignored by the semantic analyser.

1.8.5 Integration

The current system uses higher level information to enhance recognition performance. However, existing higher level processes only have a limited opportunity to contribute information. The system uses a dictionary and lexical and contextual analysis to improve the performance of the script recognition process. Different sources of knowledge (lexical, syntactic, semantic and pragmatic) are utilized in this process. Specifically the system employs a hierarchy of contexts for different levels of information (for strokes, letters, words and sentences).

In the current system meta-word information is only used to filter, or constrain, the set of word alternatives generated by the pattern recognizer. The syntactic and semantic analysers are used to filter the set of candidates in order to make a final word selection. The higher level processes do not contribute to the set of possible word alternatives, but are simply allowed to re-order them. In contrast, contextual cues are used to contribute to word recognition within the human reading process.

The semantic processor and the syntactic processor work in isolation. The reason for this is that both processors provide the same kind of information. Both processors calculate the probability of a word being present in a given environment, one based on the word's syntactic context, and the other based on its semantic context. In other words, each of the processors attempts to determine how appropriate a word is to a given local environment. The information which one processor provides is therefore of no use to the analysis of the other. Since the two processors are unable to provide one another with assistance, they operate in parallel.

Integration of the sources of information is very crude. A simple combination of the scores generated by each of the processors is used to determine the final choice of the system. The semantic and syntactic processors simply order the list given to them by the pattern recognition system. A simple voting procedure is then used to select the final choice [Rose, 1992]. The syntactic analyser is more often correct, but produces a greater number of tied results. If both processors suggest the same word then it is chosen, otherwise the choice from the syntactic analyser is used.

It has been stated that a robust script recognition system requires the use of context. The reason for this is the ambiguity of cursive handwriting. Pattern recognition must therefore be allied with other sources of information if the identity of ambiguous input is to be resolved. However, no single source of information is flawless. Integrating different sources of information will improve efficiency, but it will also make the machine system more robust. It is necessary to use several sources of information because natural language is inherently ambiguous. Natural language displays syntactic ambiguity, e.g. the choice between adjective and preposition in the sentence: "the chickens are ready to eat", as well as lexical ambiguity, e.g. 'nurse' could be a noun phrase or a verb, as well as semantic ambiguity, e.g. synonyms. A robust system must therefore include several different sources of information. There is still no guarantee that it will always be possible to disambiguate input. However, the greater the number of sources of information available to a machine system, the more likely it is that at least one of the sources of information will be capable of disambiguating confusing input.

The contribution of meta-word information to the recognition system is considered in Chapter 6. Chapter 6 looks at the ways diverse sources of information can be combined and examines ways in which an integrated system can be further developed. However, throughout the thesis, situations in which it is felt that meta-word information will be of help in the recognition process will be indicated.

1.9 Summary

The recognition of cursive script is problematic because handwriting displays considerable variation in letter form and because characters have ambiguous boundaries. The existence of ambiguity means that a machine system which relies solely upon pattern recognition will not be as accurate as one which can draw upon other sources of information as well. There are limits to the recognition of isolated letters, i.e. letters without any supporting contextual information. Contextual cues at the word level and the meta-word level influence the performance of human readers. It should be possible for these cues to be applied and integrated into the recognition process in order to improve machine performance. Script recognition systems thus require the use of context to disambiguate input efficiently. Contextual cues provide a significant advantage to the human reader in the successful recognition of cursive script. These cues can be applied and integrated into the recognition process in order to improve machine performance. An examination of the way in which humans read is not intended to place any limits on the sources of information used by a machine system nor upon the kind of methods adopted. Rather such an examination will be used to expand the sources of information that can be drawn upon by a machine system.

One characteristic of cursive handwriting, therefore, is ambiguity. However, cursive handwriting also shows a high degree of redundancy. The existence of redundancy means that many different sources of information can be utilized by a machine system for the purpose of script recognition. Furthermore, it is possible to use sources of information which human readers do not, or only use in a limited fashion. It is also reasonable for a machine system to use information in a way which is different from a human reader. Machine systems which incorporate both pattern recognition and sources of meta-word information have already been created. The development of methods for applying word level cues will lead to a further improvement in machine performance.

1.10 Outline of the Thesis

Chapter 2 reviews research into human reading and cursive handwriting recognition. This review will, firstly, provide the reader with background information about the use of word level cues by human readers. Psychological research about the human reading process will be reviewed. Human reading models presented in the psychological literature will be described. Secondly, Chapter 2 will review research into the computer recognition of cursive script. In particular, this review will describe the use of word level cues in handwriting recognition systems. Lastly, a justification for the word level cues used by the word level method will be constructed.

Chapter 3 presents an experiment about the Word Superiority Effect. This experiment will confirm its presence in the reading of cursive handwriting by human readers. A comparison will be made between the performance of human readers and the pattern recognizer on the same sample of cursive script. Human readers demonstrate an improvement in performance between letters and word. This will be contrasted to the performance of the machine system. It will be argued that the performance of the machine system can be improved if it utilizes word level cues.

Having established that word level cues are expected to lead to an improvement in machine performance, Chapter 4 presents an initial exploration of a method to use these cues. The method of applying the word level cues will be called the word level method. Firstly, a way of deriving word level cues from the list of candidate words generated by the pattern recognizer is described. Word level cues are subsequently used to derive a list of word candidates. Chapter 4 will demonstrate that the word level method can be effective and that it can be successfully integrated with the pattern recognizer.

Chapter 5 describes ways to improve cue derivation. The method of deriving word level cues from the candidate list is expanded and improved upon. An alternative, and perhaps more natural, source for these cues is pattern recognition. A method, called direct cue extraction, to recognize the cues from the input data is described.

Chapter 6 describes, firstly, the development of an improved version of the word level method and presents, secondly, an evaluation of the developed method. The way in which word level cues are applied is re-examined and improved upon. The method presented in Chapter 6 differs significantly from the method described in Chapter 4. The use of word frequency information is investigated. The integration of the word level method with the pattern recognizer is reconsidered. A way to apply letter verification procedures to the integrated system is described.

The second part of Chapter 6 is an evaluation of the word level method. One objective of the word level method is the creation of a system suitable for post-processing. The extent to which the word level method has met this objective is examined. The differences between the word level method and a word shape recognizer are examined. This examination includes a comparison between the performance of the word level method and a word shape recognizer. Lastly, the relevance of the word level method to the word superiority effect is considered.

Chapter 7 discusses the work presented in the thesis. All of the methods presented in the thesis are reviewed. It is concluded that the word level method has been a success. Integration of the word level method with the pattern recognizer produces an improvement in machine performance in all respects. Further work arising out of the research is described.

Chapter 2: Literature Review

2.1 Introduction

This chapter reviews research into human reading and cursive handwriting recognition. This chapter will provide

An examination of research into the use of word level cues by human readers.

A description of the use of word level information in handwriting recognition systems.

A justification for the word level cues used by the word level method.

2.2 Human Readers

2.2.1 Introduction

There is a great deal of evidence in the psychology literature for the use of context by human readers. In this chapter literature which examines the human reading process is reviewed. The objective of this review is to identify what kinds of word level information are available and useful to the recognition of cursive script. It is possible to look at the human reader in order to identify the sources of information used in the human reading process. Ways to improve machine performance can be suggested by such an examination of human recognition. The expertise of human readers is a natural starting point for studying factors which can be used to improve the performance of a machine system. For example, the ambiguity of cursive script means that it is not always possible to recognize cursive handwriting by pattern recognition alone. However, human readers manage to read even unfamiliar, degraded or ill-formed handwriting quite successfully. It is evident that human readers are using other sources of information alongside pattern recognition. An understanding of human readers will identify, firstly, the various sources of information exploited by human readers, secondly the reliance which human readers place upon these different sources of information, thirdly the relative importance attached to the different sources of information, and lastly the way in which different sources of information are combined.

For instance, eye movement studies provide evidence of the redundancy present in English. Human readers direct their reading according to information obtained in parafoveal vision, as well as in foveal vision. A number of studies show that the information gained in parafoveal vision is used by human readers to aid word recognition and, in some circumstances, to identify a word without recourse to further examination. The kind of information which can be gained in parafoveal vision therefore provides important evidence about the human reading process and the particular features used by human readers to recognize words. Parafoveal studies, firstly, indicate the kind of information which a human reader can use. It is possible to recognize words merely using parafoveal vision. Parafoveal vision is not as clear as foveal vision. The information gained by parafoveal vision is less detailed than the information gained by foveal vision. Parafoveal studies therefore also indicate that there exist features which are so informative that they can be used to identify a word without further processing. This implies that word level cues can be used to recognize a word. If it is possible to identify a word using parafoveal vision then this suggests that there are sources of information being used by human readers which a machine system should be able also to exploit. Parafoveal vision is used to direct attention. Parafoveal studies provide evidence about the integration of, and interaction between, different sources of information, e.g. it is possible to show that there is an interaction between contextual constraints and parafoveal visual information [Balota & Rayner, 1991]. This can help show, for instance, the point at which a particular source of information is used within the human reading process.

2.2.2 Eye Movements

It has been suggested that examination of eye movements can provide evidence of how text is being processed by the reader. The eye does not travel smoothly across text during reading. The eye makes abrupt movements, or "saccades", between fixations. An average of 90% of reading time is spent in fixations. Readers tend to fixate near the middle of words. Longer words are more likely to be fixated than shorter words and they tend to be fixated for a longer period. Roughly 80% of content words, and 40% of function words are fixated on in normal printed text. Fixations also become longer and increase in number where the text is more complex or difficult, e.g. containing longer or more technical words, or dealing with difficult concepts. The average fixation duration is approximately 225 milli-seconds. However, the visual cues necessary for reading can be obtained within the first 50 milli-seconds or so of a fixation.

Movement between fixations takes, on average, about 20-40 milli-seconds. Saccades are faster within a line of text. The movement of the eye from the end of a line to the start of the next line takes approximately 40 milli-seconds. The mean length of a saccade when reading printed text is about eight character spaces. The length of the word immediately to the right of fixation influences saccade length. The longer the word, the further the eye tends to jump. Between 15 and 20 percent of saccades are regressions, where the reader makes a right-to-left saccade back from the current point of fixation. Approximately 5-20 percent of content words receive more than one fixation [Rayner & Balota, 1989].

The reading speed of an adult is about 250-300 words per minute. 300 words per minute being characteristic of rather easy, non-technical text [Barber, 1988]. The processes involved in reading are typically very fast. The reader is unconscious of the vast majority of these processes. For example, text on a computer screen was moved a short distance to the right or left during certain saccades [O'Regan, 1981]. Although this shift tended to cause a corrective eye movement, the subjects were unaware that the text had been shifted

It is more likely for a predictable word to be skipped than an unpredictable one. Studies about the probability of fixating a target word and the duration of fixation provide evidence that readers use predictions from syntactic and semantic constraints. The different fixation rates for function and content words show the influence of contextual constraints. Fixation of content words can be shown to be influenced by semantic considerations. A target word in a highly constrained context is less likely to be fixated than the same word in a poorly constrained context. Predictable target words were also fixated for a shorter time than unpredictable ones [Ehrlich & Rayner, 1981].

2.2.3 Foveal and Parafoveal Vision

The very centre of the eye is called the "fovea". A distinction can be made between the information received in foveal and parafoveal vision. Visual acuity is strongest at the fovea. The central parts of the image around which the eye is fixating are received by the fovea. The word currently fixated will therefore be seen most sharply by the eye. The central two degrees of vision around the point of fixation fall on the fovea. However, information about the material being read is also gained in parafoveal vision. The parafoveal region extends beyond the foveal region, up to 5 degrees of vision to the left of fixation and to the right of fixation. Lastly, there is a region of peripheral vision which extends beyond the parafoveal region out to the range of vision. Balota and Rayner suggest that in normal reading conditions three or four letter spaces of printed text subtend one degree of visual angle. The fovea would approximately include six letter spaces around fixation. The parafovea would include the next twelve letter spaces to the left and right. The periphery would include information beyond fifteen letters from fixation. The distinction between foveal, parafoveal and peripheral vision is based on the physiological structure of the retina [Balota & Rayner, 1991].

The perceptual span (the region of effective vision) therefore extends from the beginning of the currently fixated word, or about 3-4 character spaces to the left of fixation, and about 15 characters to the right of fixation. The perceptual span is asymmetric to the right in English. Different kinds of information appear to be available at different distances to the right of fixation. The area closest to fixation and extending to 4-8 character spaces to the right of fixation (foveal and beginning of parafoveal vision) obtains information used to identify the word in the current fixation. The region in which words are identified is variable since on some fixations one word can be identified, whereas on others two or more can be seen depending on whether short words occur together. Further to the right of fixation than the region of word identification, beginning-letter and some letter-feature information is extracted. Word length information appears to be acquired over the largest range (out to about 15 character spaces) [Rayner, et. al., 1982; Rayner & Balota, 1989]. Although the perceptual span is asymmetric to right in English, it is asymmetric to the left in Hebrew (Hebrew is read from right to left).

It can be argued that the kind of information which can be extracted in parafoveal vision may give some indication of human reading processes. It should also show the importance which human readers place upon different sources of information and so indicate the main visual cues which a reader uses. There are a number of reasons why this is thought to be so.

Human readers appear to gather only as much information as is needed to identify a word. They cease extraction of further visual information once an activation threshold is reached and an appropriate word has been suggested.

> ... the use of only a fraction of the available cues represents the normal rather than the abnormal mode of recognition. The more perceptual evidence is required, the less efficient use is made of implicit knowledge or expectations of word forms, grammatical structures, meaning etc. Efficient reading would therefore seem to require a restricted and flexible use of perceptual information, just sufficient for reconstructing the words of the passage or for understanding its meaning. [Bouma, 1971]

Words seen outside foveal vision are sometimes skipped over during reading (e.g. the eye tends to skip over short function words), but are successfully identified by the reader [O'Regan, 1979].

It can be shown that parafoveal vision improves performance. The parafoveal preview effect is well established. The presentation of a target word to the parafovea facilitates identification of that word [Dodge, 1906]. For example, restricting human perception of text to the word fixated reduces reading speed to about 60% of the normal rate. This implies that "significantly more information than the fixated word is extracted on many fixations" [Rayner, et. al., 1982].

One of the main experimental methods used to study parafoveal vision is parafoveal orthographic priming. A word is presented to the subject's parafovea.

During the saccade, the word is replaced by a target word that is identical, orthographically similar or orthographically dissimilar to the preview word. Similarity between the target word and the preview word facilitates identification of the target. [Rayner, et. al., 1980; Balota & K. Rayner, 1991]. This shows that there are important cues which are recognized in parafoveal vision and which are used to aid recognition performance.

2.2.4 The Word Superiority Effect

Human readers show a strong context effect at the word level. This is known as the Word Superiority Effect (WSE) [Cattell, 1886; Baron & Thurston, 1973]. This is the effect noticeable with human readers whereby letters in words are recognized more easily than letters in isolation or letters in non-words. This effect is well documented although no explanation for the effect has been agreed upon [Monsell, 1991]. Generally, some knowledge of how letters combine to form words aids their recognition in some way. The WSE shows that human readers utilize contextual cues to increase their recognition at the word level.

The WSE can be considered a low level context effect. Reading performance is better on words than performance on single letters. It could be argued that each letter of a word acts as a context for the others. However, there cannot be a general effect just from the context of other letters.

It is easy to demonstrate the effects of word level context upon perception. For instance, it is less easy to read and remember letters if they appear as a meaningless string, e.g. the non-word "eeopdvl", than if the same letters appear in an more conventional order, e.g. the pseudo-word "vedolep", and this effect is even more pronounced when the letters are rearranged into a meaningful word, e.g. "develop".

The reason why human readers find it easier to read whole words rather than isolated letters lies in their ability to take into account contextual information at the word level. It is rare for human readers to become hopelessly confused even with less clear writing. Reading interweaves perceptual and cognitive processes. Human readers are obviously guided by some knowledge about what is likely and what is possible. This appears related to knowledge about probable letter confusions and other types of word shape information, but it is also related to word frequency and lexical knowledge, although in this latter case, it may be a case of conforming to English spelling regularities rather than straight forward lexical lookup. It is certainly the case that given some sort of context human readers use this to guide their recognition of characters and words.

A number of factors which affect human recognition at the word level, have been identified.

lexical information

word frequency

orthographic regularity

phonological regularity

morphological information

word shape

Each of these factors will be considered in turn, and their relevance to the machine recognition of cursive script examined. Human readers use word level cues as an important aid to their recognition of words. A machine system that does not exploit contextual cues cannot be expected to show the WSE.

2.2.5 Lexical Information

One cause for the WSE appears to be lexical information, i.e. choosing a letter that completes a word. However, lexical constraints are not sufficient by themselves to produce the effect. Two response alternatives may both make up an acceptable word (e.g. "cat", "cot"), but more often than not a human reader will make the correct response. A machine system which attempts to exploit word level cues should use lexical information. Lexical information is already used by the machine system since letter strings are filtered according to a lexicon to produce a set of word candidates. Lexical information provides the set of English words known to the system. It determines the letter sequences which are allowable.

2.2.6 Word Frequency

One factor which can be shown to make a contribution to the WSE is that of word frequency: human readers tend to recognize higher frequency words better than lower frequency words [e.g. Broadbent, 1967]. Incidentally, the reason why most readers will tend to interpret the word in Figure 1-7 as "dog" rather than "clog" or "cloy" is that the first of these words has a much higher word frequency than the others. The word frequency effect is well documented. High frequency words produce better performance than low frequency words across a wide range of word recognition tasks. There is some evidence that frequency also influences the use of parafoveal information [Inhoff & Rayner, 1986]. Visual duration thresholds for words are a function of word frequency [Howes & Solomon, 1951].

The locus of the word frequency effect is the subject of some debate [Monsell, 1991]. A number of factors may be involved. Factors co-vary with frequency (such as orthographic regularity and phonological regularity) and have also been shown to influence recognition performance, independently of lexical status [Tannenhaus, et. al., 1980]. Different theories of human word recognition propose different sites for this effect. The word frequency effect could be a reflection of some property of the perceptual mechanism, or could be a response bias from the subjects' greater tendency to use high-frequency words.

The Logogen Model [Morton, 1969] proposes recognition units, called logogens, which collect evidence from a number of sources (physical, syntactic, semantic, cognitive) for the presence of a word. Once a logogen has collected sufficient evidence it is said to "fire", thus recognizing a word. The more evidence accumulated which is consistent with a particular word, the more its logogen will be activated. Morton proposes that logogens are tuned for word frequency. Logogens for higher frequency words have lower response thresholds for recognition than low frequency words. The effect of this is that high frequency words exceed threshold sooner than low frequency words and therefore are available for response more quickly. The effects of context were also incorporated by Morton. The logogen system is assumed to receive input information from many sources. The recognition threshold of a logogen is temporarily lowered by context. Previous semantic and syntactic input will serve to lower the threshold of words that are related to or belong in the stimulus context. The threshold of a logogen which is consistent with previous semantic and syntactic input will therefore be reached more rapidly than the threshold of one which is not. Successful recognition also lowers a logogen's threshold. This accounts for repetition effects. Repetition causes a long lasting, but slowly decaying, reduction in the threshold of the logogen.

McClelland and Rumelhart proposed that there is interaction between the word and letter levels in recognition [McClelland & Rumelhart, 1981; Rumelhart & McClelland, 1982]. McClelland and Rumelhart's interactive activation model uses forward-going activation between levels and competitive inhibition within levels. Word level units can feed back to letter level units, either exciting or inhibiting them, and letter level units can excite or inhibit word level units. Letter level units collect information about the presence of letters. The input activates units at the letter level, which in turn feed information to word level units. The words which receive the greatest amount of activation from the letter level are those that are most orthographically akin to the input. Each word which has been activated competes via mutual inhibition with all other activated words. Words therefore facilitate the recognition of their constituent letters by activation being fed back from the word-percept level to the letter level thus improving the perception of individual letters [McClelland, et. al., 1992]. Word frequency effects are the result of higher frequency words having higher baseline activity than lower frequency words.

Forster proposed that words are initially selected on the basis of their physical features; candidate sets are then narrowed down and ordered on the basis of their frequency [Forster, 1976]. One general explanation for frequency effects is, therefore, that human readers exploit visual cues together with frequency, contextual and lexical information to derive an expected word. This appears to suggest that human readers derive only enough evidence from their visual examination of a given word to suggest a most likely candidate. As Broadbent points out:

If the subjects were biased in such a way as to accept a smaller amount of evidence before deciding in favour of a probable word, the word frequency effect would be obtained [Broadbent, 1967].

It can be argued that the word frequency effect suggests that humans place limitations on the lexical search space.

Word frequency does play a part in the WSE. However, word frequency has not been explicitly used by those involved in the machine recognition of cursive script, except in the limited role of selecting a lexicon. This is in marked contrast to the approach taken by such as McClelland and Rumelhart, where frequency takes a central role in the computer systems which they have developed. Since word frequency can be shown to play a part in reading, it should be possible to utilize it as an effective part of the machine recognition of script. Word frequency can be used to decrease the size of the lexicon which is used at any one point. This is useful if the recognition method is not discriminatory. Word frequency can also be used to select a word in those cases where all other methods of recognition or selection have failed. Word frequency information will therefore be used as part of the word level method.

2.2.7 Orthographic Regularity, Phonological Regularity and Morphological Information

It is difficult to disentangle the effect of orthographic regularity from that of phonological regularity since pronounceable letter strings usually conform to English spelling regularities. The use of morphological information by human readers is also connected with the possible use of sub-word identification. These three sources of information will be examined together.

Orthographic regularity is used to facilitate letter identification, e.g. the use of letter bundles or sequences. Orthographic regularity has been measured in three related ways: the frequency of component letters and sub-strings [e.g. Mason, 1975], the exploitation of statistical redundancy [e.g. Thompson & Massaro, 1973], and the use of spelling regularities [e.g. Massaro, et. al., 1979]. The high degree of redundancy in written English has been discussed in Chapter 1. It is true that good readers find it easier to encode (e.g. to memorize and to recognize) letter strings which conform to English spelling regularities than arbitrary strings of letters [e.g. Baron, & Thurston, 1973].

Phonologic coding does appear to play a role in word identification [e.g. Humphreys, Evett & Taylor, 1982]. English does not have one letter for each phoneme. Furthermore, of the 26 letters of the alphabet, three letters are superfluous to the task of representing phonemes (c, q, and x). In all, 23 letters can be used to represent the 44 phonemes used in English. The letters of pseudo-words that can be pronounced and which follow the conventions of English spelling are recognized more easily by human readers than the same letters in non-word strings (e.g. "lant" compared to "tnla") [Just & Carpenter, 1987].

There appears to be little evidence that morphemic information is used. For instance, human readers do not appear to make use of the beginning morpheme of a word, e.g. identification of a prefix [Lima, 1987]. However, there is some evidence that previewing the first morpheme of a compound word (e.g. 'cow-' in 'cowboy') facilitates processing of the target word [Inhoff, 1987].

There is no strong evidence that orthographic regularity, phonological regularity and morphemic coding are used in parafoveal identification. Since the use of parafoveal vision involves integrating information across saccades then it indicates at what point a particular source of information is used during the human reading process. Rayner and Pollatsek observe that

> ...words and abstract letters are likely to be active units in integration across saccades in reading or word identification, while sound codes and letter features are not. The only evidence of "deeper" subword units being active in integration is that the first morpheme of a compound word (a word itself) appears to be a unit. [Rayner & Pollatsek, 1987]

It can be suggested from this that for words (in comparison to pseudo-words or even non-words) orthographic and phonological effects are post-lexical, i.e. they occur after lexical information has been accessed, or at least concurrent with the use of lexical information.

There are three ways of treating lexical information at the word level in a machine system: the use of a fixed dictionary and matching words to this dictionary, the use of statistical information about transition probabilities between characters, and a hybrid approach incorporating elements from both of these techniques. [Ford & Higgins, 1990]. The use of statistical information about transition probabilities will generate the most probable word but does not guarantee that the word is valid. This was one reason for the development of the hybrid approach since this approach incorporates a dictionary search. Whether the dictionary matching approach or the hybrid approach is used is irrelevant to the purpose of obtaining lexically valid words. The only considerations relevant to the approach adopted are how far allowance is made for incorrect character segmentation, the retention of potentially useful character and segmentation information, speed of access and processing, size of memory required, and the ease to which wild card substitution for characters can be implemented. However, the hybrid approach could also generate words which are not in the dictionary but which are deemed to be likely.

The information which orthographic and phonological regularity can supply is inherent in the use of a lexicon. It makes no difference to the outcome whether the statistical properties of English orthography are used to determine that the letter 'u' is more likely to complete the letter sequence 'eq', or whether an incremental method of matching letter strings to known letter patterns (the method adopted in multiple interactive segmentation) is used. This lack of a difference is true, however, if, and only if, what is desired is the identification of words present in a lexicon.

Without orthographic regularity, the type of errors caused by misidentification will not be the same for machine as for human. The machine will not be able to identify pseudo-words with the same facility as human readers. A system which uses a comprehensive set of probabilistic rules will be able to deal with whatever word it is presented with. In other words, it is not limited to a given lexicon. The ability to treat any word can be contrasted with the existing lexicon based system. Given that new words are continually introduced into the language, no lexicon can be complete. Verification procedures can use letter sequences. It is easier, in this situation, to use information about letter sequences or English spelling regularities than it is to search a lexicon. This is an implementation issue. The use of orthographic information for word verification is examined in Chapter 6. The current system treats words as discrete units. It may be the case that the first morpheme of a compound word should be used as well.

Orthographic regularity, phonological regularity and morphological information will not be used in this present work. It is not agreed that these three sources of information do play a part in the WSE. The introduction of these factors into the recognition process is not a simple task. For example, the machine system must still be able to recognize irregular words when they occur.

2.2.8 Word Shape

Human readers obviously use visual cues to identify words. However, exactly what kind of cues are used, and their relative importance, is the subject of much debate. In proofreading studies it has been shown that misspellings that do not change the overall shape of a word are less likely to be detected than ones that do [Haber & Schindler, 1981]. This is a comparison between same and different shape letter substitution.

A high level approach to recognition is the identification of a word as a whole by its shape. Some writers have argued that the outline (or envelope) of the shape of the word is used to identify words [e.g. Cattel, 1886; Crowder, 1982]. However, there is no evidence which unambiguously supports the argument that words are wholly, or in part, identified by their particular outline shape [e.g. Johnston, 1981; Besner & Johnston, 1989]. A less generalized form of the outline shape approach holds that it is only short high frequency words which are identified in this way. For instance, Taylor and Taylor suggest that

Contour is used, at least in reading short words. It is especially important for function words. In many cases, people identify the function word only by its outer contour, even when the readers are looking for misprints. [Taylor & Taylor, 1983]

Word shape cues, or transletter word shape cues, cannot be the main way of identifying words, if they are used at all, since words displayed in formats that destroy these cues can still be read reasonably well. This can be seen from the results of case alternation experiments (e.g. aLtErNaTiOn, AlTeRnAtIoN). The destruction of visual cues by case alternation can, however, be sometimes overestimated. Case alternation preserves a reasonable amount of word shape.

For instance, word length is unaffected. Secondly, some letters are have very similar upper and lower case forms, e.g. c-C. The most important point is that the identification of actual letters (i.e. their identity rather than their visual form) is unaffected.

2.2.8.1 Abstract Encoding of Letters

Case alternation also suggests that letter information is used at the abstract letter level and is not visually based. The meaning of abstract encoding is that, for human readers, characters are represented internally using a letter coding. It is the identity of the letter which is encoded, not its visual form. The effects of parafoveal previews are relatively indifferent to changes in case between the parafoveal and the foveal word [McConkie & Zola, 1979]. The effect of priming is not solely due to visual similarity between primes and targets. The active units in word identification are letters rather than letter features [Rayner & Pollatsek, 1987].

However, parafoveal orthographic priming experiments indicate that letter information is used by human readers, e.g. identification of the target word is easier when the preview word resembles the target (see section 2.2.3). It may not be the characters as written on the page which cause letter confusions or cause the preservation of shape, but the abstract letter coding [Humphreys, Evett & Quinlan, 1990].

Encoding to abstract representation occurs very quickly in reading. Characters are not stored using their visual pattern, but are rapidly transformed into abstract mental representations. Furthermore, the priming effects of middle letters are position independent: priming occurs but it is not affected by the specific position of middle letters [Humphreys, Evett & Quinlan, 1990]. Lastly, letter confusions tend to share common properties of shape with the intended letter [Bouma, 1971]. It was therefore decided that an abstract representation of letters would be relevant to the machine recognition of cursive handwriting at the word level. This representation is a set of generalized cues which preserve the sort of information retained across letter confusions. These cues are the presence or absence of ascenders, descenders, i-dots and j-dots, and lastly t-crosses and f-crosses (see section 2.4).

2.2.8.2 First and Last Letter

It has been suggested that identification of just the first and last letters of a word can, in some cases, be enough to identify the word [e.g. Taylor & Taylor, 1983]. The first character of a word tends to be written more clearly than succeeding ones. The difficulty of segmentation is also slightly eased in the case of the first character of a word since it lacks any preceding character. Segmentation difficulties are also eased in the case of the last letter of a word. However, the last letter of a word often tends to be written less clearly than the other letters. First and last letters also tend to provide more information about the word than central letters. The central letters of a word tend to be vowels, whereas the first and last letters of a word tend not to be. It has been suggested that the priming effect of middle letters is not affected by their specific position. End letters, however, tend to produce stronger priming effects than middle letters. During word recognition the specific positions of middle letters are not coded. However, the position of letters as internal or end-letters do appear to be coded. [Humphreys, Evett & Quinlan, 1990]. The reading frameworks presented by McClelland and by Mozer lend support to the argument that end letters are more important than middle letters [McClelland, 1986; Mozer, 1987]. Both of these frameworks use a position independent coding of characters. For instance, McClelland used a coarse coding of letter representations in which letters were coded in terms of whether there was a blank or filled space to the left or to the right of the letter. These frameworks suggest that first and last characters are most significant to recognition. This shows that first and last letters are relevant to the machine recognition of handwriting at the word level.

The existing pattern recognition system places the same reliance upon all of the letters of a word, irrespective of their position. Letter segmentation and character matching difficulties mean that the first character of a word may still be ambiguous. For instance, initial characters are sometimes written with a leading line and it still remains the case that an initial character has to be distinguished from the immediately succeeding character. There is always the possibility that a single character may be interpreted as two characters, and vice versa. A machine system can only place a greater reliance upon the first character (in comparison with later characters in the character string), and not absolute assurance.

2.2.8.3 Word Length

The fact that human readers use word length as an aid to recognition at the word level is well established. Human readers can identify word length in parafoveal vision, and there may also be possible identification of length in peripheral vision. Word confusions tend to preserve length. It has been suggested that parafoveal information about word length plays no part in word identification, but is only used to determine the next fixation point [e.g. McConkie, et. al., 1982]. This is convincingly dismissed by Balota and Rayner [Balota & Rayner, 1991]. It was therefore decided to use word length (measured by the number of characters) as one of the cues in the word level method.

2.3 *Word Level Information in Script Recognition Systems*

2.3.1 *Introduction*

Initial attempts at the machine recognition of handwriting date from the early 60's [e.g. Frishkopf & Harmon, 1961; Earnest, 1962]. A great deal of research effort has been expended on the topic. Handwriting has been seen as a viable, and natural, alternative to the keyboard (see section 1.2). A large number of possible applications for cursive script recognition exist [e.g. see Higgins & Ford, 1991].

A number of divisions exist within script recognition. Handwriting recognition applications are either on-line or off-line. On-line recognition systems use a special input device (such as electronic pen and paper). In contrast, the input to off-line recognition systems is a digitalized image. On-line systems have access to temporal information. This is useful because the main body of a word tends to be written in sequence from left to right, although a writer may go back to add dots and crosses later. It is also possible to use penlifts to indicate the beginning and end of words and temporal information can also be used to determine upstrokes and downstrokes. There has been research into recovering dynamic information from static data [e.g. Boccignone, et. al., 1993]. This would allow off-line recognition systems access to this valuable temporal information.

Broadly speaking, there are two main approaches to the problem of recognizing cursive script: word shape recognition, and segmentation followed by letter or stroke recognition. The emphasis of this review will be upon word level, typically word shape, handwriting recognition, because this approach is most relevant to this thesis. Word shape recognition has also been called whole word recognition, or holistic recognition. Word shape recognition, as the name suggests, treats the word to be recognized as a whole, and not as a sequence of characters. The features of a complete word are obtained and matched against a database of stored templates in order to find the closest match [e.g. Brown & Granapathy, 1980]. The advantage that this approach has over the segmentation approach is that the difficulties of segmentation are bypassed. However, such a system may be restricted to a small lexicon since the system must be explicitly trained for every word in its vocabulary.

The use of word level cues has been neglected in comparison to the segmentation approach because it is often felt that they are insufficient to recognize a word unambiguously [e.g. see Madhvanath, et. al., 1997]. This view is challenged below (see section 4.2). Word shape recognition has been more commonly applied to off-line recognition, than on-line.

The alternative to word shape recognition is to treat words as sequences of smaller size units. These units can be individual characters or strokes [e.g. Higgins & Whitrow, 1984]. Recognition in this approach relies on the input data being segmented into the segmentation units. The segmentation units are then matched against a database of known patterns and the closest match selected. The recognizer hence builds up words from their individual units.

An intermediate approach between word shape recognition and segmentation recognition is the use of lexical information. Contextual information is often understood as lexical information only [e.g. Elliman, 1990]. The lexical entries that match the character strings obtained from character recognition are selected. It is possible to employ methods of selection in which the necessity to recognize each and every letter of a word is relaxed. For example, string correction algorithms can be employed. This enables the selection of entries that best match the character strings, although one or more characters may be missing or wrongly identified. Wildcards can be used. Wildcards allow characters which have not been explicitly identified by the recognizer to be considered as present in the word.

Four different uses of word shape information will be examined:

- word shape recognition as the primary recognition method
- word shape analysis used to reduce the size of the lexicon
- the use of lexical information by segmentation recognizers
- the combination of a top-down method with verification techniques

Lastly, the integration of knowledge sources will be discussed.

2.3.2 Word Shape Recognition

An early recognizer that used word level information was that of Frishkopf and Harmon [Frishkopf & Harmon, 1961]. Letters were recognized by means of a binary decision tree using features such as retrograde strokes, cusps, and closures. The word was also separated into three horizontal zones. Special features (e.g. ascenders, descenders and retrograde strokes) were used to locate stroke segments. A 100 word dictionary was used.

A word recognizer was developed by Earnest at the start of the sixties [Earnest, 1962]. Whilst data was captured electronically using a electronic pen and cathode ray tube, it was subsequently analysed using a binary matrix and therefore the system could be considered to be an off-line recognition system. Feature extraction involved, firstly, an estimate of the envelope of the central letters of the word (i.e. a determination of a horizontal area which did not contain any ascenders or descenders). Secondly, key features of the word were found. These features were ascenders, descenders, crosses, closed curves, and the number of times the horizontal midpoint of the word was crossed. A code was constructed using these features and a dictionary searched for words which matched the generated code. A 10,000 word dictionary was used and Earnest reported that 18% of the words in a test sample were uniquely identified by the system.

In the approach outlined by Miller, the input word was decomposed into macro-feature segments which, it was argued, made up the set of basic writing units [Miller, 1971]. Miller suggested that the shape of any word could be described using a combination of these units. The units had been derived using shape analysis. The units were not necessarily related to characters, e.g. a single unit could be part of several characters. A unit was coded to a value representing its shape. Previously obtained values held in a lexicon were compared to the input data to obtain the most likely word.

A Markov model is a combination of states that are connected by transition probabilities. One of the first researchers to employ Markov Models in the field of cursive script recognition was Farag [Farag, 1979]. Farag's system attempted recognition on a word-level basis. A word was treated as a sequence of directional strokes. The model of representation was a nonstationary Markov chain, with the states of the Markov chain being the strokes. Each word was represented by a collection of stochastic forward transition matrices. It was intended that only a small number of key words would be classified by the system. The model truncated longer words which meant that, for example, it was unable to distinguish between the words "class" and "classify".

Features used in the system developed by Brown & Granapathy were upper and lower loops, word length in characters, ascenders, descenders, dots, crosses, cusps, retrograde strokes, closures (e.g. 'b' possesses the closure feature whilst 'c' does not), and openings (including the direction of the opening, e.g. 'c' has an opening to the right) [Brown & Granapathy, 1980]. A feature vector was formed based on the number of occurrences of each type of feature. The feature vector was matched against vectors stored from a training phase to find the nearest match. The system had a small recognition vocabulary.

The rationale behind word shape analysis for Ho and colleagues was the recognition of word images that were degraded and therefore especially prone to errors in character segmentation [Ho, et. al, 1992a]. The input word was first partitioned into a fixed grid. This grid provided a global frame of reference and was used to represent the locations of shape features. The word was divided by four reference lines into three horizontal regions: the ascender region, the middle region and the descender region. Many characters are located entirely in the middle region. To facilitate more accurate position descriptions, the middle region was further divided into upper and lower parts. The shape of a word was described using a set of features referred to as the stroke direction distribution. Each word was represented as a feature vector. The vector was matched against a lexicon of words and a ranked candidate list produced. A variation on Edit Distance was used to match input and prototype (see section 2.3.4).

Simon argues for an approach based on first finding regular features, and then finding singular features [Simon, 1992]. Anchor points (features which are reliable) were used to select a list of candidate words from a lexicon. A large number of contextual rules were used to distinguish between words. A subsequent examination of the candidate words was carried out to see if there other letters matched less reliable features. A small lexicon of 25 words was used.

Knowledge sources used in the system developed by Higgins & Bramall include the overall shape of the word, a word shape thinner, a word shape predictor, a downstroke detector, and a turning point detector [Higgins & Bramall, 1993; Higgins & Bramall, 1994]. The input is tested against a template for each word in the lexicon and the best match selected. A blackboard control model is used in the system (see section 2.3.6). A training phase allows the performance of the feature detectors to be analysed. This analysis is subsequently used to decide which knowledge source to implement and to assist in the scheduling of the knowledge sources. Knowledge sources are themselves responsible for indicating to the system how much of a contribution to the solution they are able to make. A knowledge source may be responsible for indicating when a satisfactory result has been found, e.g. when the activation value of one word crosses a threshold. Higgins & Bramall have drawn upon the logogen model of lexical access (see section 2.2.6). A list is generated of the most likely candidates. Particular differences between candidates words are then identified. This information is used to decide which of the candidates best matches the written word.

A whole word recognizer has been developed at the Nottingham Trent University [Powalka, et. al., 1994; Powalka, 1995]. Powalka calls this a "wholistic recognizer". This recognizer exploits word shape information. This includes zoning information, which is used as a guide for locating the number and location of vertical bars, word length, Word shape information is combined with independent letter verification procedures. The word shape recognizer uses the physical characteristics of the input but bypasses the exacting requirement of identifying all of the characters of a word in favour of recognizing the overall shape of the word and subsequently attempting to verify individual characters in order to produce a 'best fit' of word to shape. It attempts to match word shape information against a database of stored shapes and then attempts to verify the existence of particular letters within the input data in order to select from the word alternatives.

2.3.3 Lexical Reduction

In all of the approaches described above, word shape information has been used to recognize the target word. In contrast, Madhvanath & Govindaraju use word shape analysis for the purpose of lexicon reduction [Madhvanath, et. al., 1997; Madhvanath & Govindaraju, 1997; Madhvanath & Govindaraju, 1998]. They report that their method is presently capable of reducing a lexicon to one-half its size with almost no error. Three 'global features' are used: ascenders, descenders and word length. The local minima on the outer contours of the word are clustered into descenders and normal minima. The local maxima are clustered into ascenders and normal maxima. The minima divide the image into "pseudo-segments". The number of such segments is used as a measure of word length. A supplementary class of features is also used. These explicitly assert specific spatial properties of the positional features, e.g. that an ascender was detected at the start of the word. The features are used to match the input with each lexicon entry. For every lexicon entry, the quality of the best match is calculated, and the lexicon ranked by quality of match. A separate segmentation based technique is used to make a final selection from the candidates proposed by the word shape recognizer [Madhvanath & Govindaraju, 1997].

Lecolinet uses a pre-recognition algorithm for detecting meaningful entities (called graphemes) [Lecolinet & Crettez, 1991]. A grapheme can correspond to one character, two characters, or part of a character. Contour analysis is used to segment the input data. An estimate of the length of the word is used to reduce the size of the lexicon. A dynamic programming method is subsequently used to match the list of graphemes with the most likely words in the lexicon.

2.3.4 Lexical Information

Lexical information is often exploited by word shape recognizers [e.g. Madhvanath & Govindaraju, 1998]. However, lexical information is also used by segmentation recognizers to verify the results of character recognition. Lexical information represents an example of word level information used by segmentation recognizers.

A number of methods have been developed which use the output from a recognition system and try to correct it. Such methods effectively acknowledge that output from a recognition system may be incorrect but still contain useful information. In one form or another, these methods draw upon lexical or orthographic information and can therefore be considered to use word level cues (see section 2.2.5 and section 2.2.7). Three main approaches to the use of lexical information can be seen: dictionary methods, Markov methods and hybrid methods.

Dictionary methods rely on dictionary look-up techniques. The input word is verified by matching it with a dictionary word. Approximate string matching can be used. It is also possible to verify whether two strings are identical or whether one is a misspelling or misrecognition of the other. Two approaches within string matching are discernible: edit distance and probabilistic methods.

Edit distance is a measure of confidence that one string is a misspelling or misrecognition of another. The edit distance defines the number of edit operations required to obtain the dictionary word from the input word [Levenshtein, 1966]. Probabilistic methods can be used to obtain probabilities that the input word is a given dictionary word [e.g. Kashyap & Oommen, 1984].

An alternative to dictionary methods is a string correction technique using the statistics of the vocabulary. An early example of this was the use of n-gram frequencies [Riseman & Ehrich, 1971]. The most common statistical technique used by string correction methods is a Markov process in which transition probabilities are assigned to various letter combinations or n-grams.

The employment of Markov Models to string correction within handwriting recognition was based on pioneering research into error correction within speech recognition [Bahl & Jelinek, 1975]. This approach to string correction can handle substitution, insertion and deletion errors as well as splitting and merging errors.

An algorithm applying Markov Models to string correction has been found to perform better than one based on generalized edit distances [Kashyap & Oommen, 1981].

Srihari & Bozinovic described a string correction algorithm tailored to cursive script recognition [Srihari & Bozinovic, 1982]. The algorithm dealt with the splitting of one letter into two, the merging of two letters into one, and the letter substitution. The algorithm attempted to obtain the best estimate of the correct word.

The Viterbi algorithm works removing or reducing unlikely letter sequences in an attempt to produce the target word [Viterbi, 1967]. A comprehensive discussion of the theory behind the algorithm can be found in [Forney, 1973]. The Viterbi algorithm deals with substitution errors only. However, improved variants of the Viterbi algorithm have been developed [e.g. Srihari, et. al., 1983].

Hybrid methods which combine dictionary methods and Markov methods have been developed [e.g. Shinghal & Toussaint, 1979; Hull, et. al., 1983; Ford &

Higgins, 1990]. Markov methods are used (e.g. Viterbi algorithm), to form probable alternatives to the input word. A dictionary is then searched. If a word is not found in the dictionary then it can be discarded. Alternatively, Markov techniques can be used to calculate the most probable result.

2.3.5 Verification Techniques

A way in which the word level method can be integrated with a bottom-up verification technique will be presented in section 6.5. There have been previous attempts to employ such verification techniques. In a system developed by Hull & Srihari, word shape features such as ascenders, descenders, and holes are used to hypothesise words for subsequent verification by a character recognizer [Hull & Srihari, 1986]. This can be considered to be a two-stage recognition system. The system created by Nadal & Suen uses a validation module alongside more conventional techniques [Nadal & Suen, 1993].

Lecolinet has proposed a top-down directed word verification method called "backward matching" [Lecolinet, 1993]. In this method, the most informative letters of the input word are recognized first and the rest of word recognized subsequently using contextual analysis. Each word in the lexicon is described by a list of letters. The letters do not follow their actual order of occurrence within the word, but rather are in a meaningful order based on the visual and lexical significance of the letters. These letters are looked for in the input word. A bottom-up feature detector recognizes primitive features such as strokes, and closed and open loops. More detailed analysis can be initiated if an expected feature has not been recognized. Backward matching combines a top-down, context driven, recognition scheme and a bottom-up, feature extraction process working in a competitive way.

2.3.6 Integration

The integration of the various different kinds of information associated with reading is an important topic in script recognition. It is certain that a machine recognition system must be able to integrate successfully diverse contextual knowledge sources about a text if it is to reach a level of efficiency comparable to human readers. The reason for this is that it is not always possible to uniquely identify a word from its written pattern. It is hoped that by integrating the different aspects of the reading process that a stronger, more robust, recognition system can be developed. In particular, one which is capable of dealing with the particular problems of unconstrained cursive handwriting and a large lexicon.

Srihari & Bozinovic describe three different models of knowledge source interaction: the hierarchical model (top-down or bottom-up), the heterarchical model (interactions in both directions, and between levels as well), and the blackboard model (connections between different knowledge sources through a global database, or blackboard) [Srihari & Bozinovic, 1987].

The recognition system which is the target of the current thesis is based on a hierarchical model (see section 1.8). The recognition system will perform better if information flows in both ways rather than just in one way. One objective of the current work is to create a system which allows communication back and forth between different levels of the system, i.e. a heterarchical system (see section 6.5).

A model which allows communication back and forth between different knowledge sources will be more efficient than one that only allows information to flow in one direction [Srihari & Bozinovic, 1987; McClelland, et. al., 1992]. Whilst a single model of the human reading process has not been agreed upon, there is strong agreement that the various aspects of the reading process interact and that the integration of diverse knowledge sources improves the recognition performance of human readers [e.g. Morton, 1969; Johnson, 1981; McClelland & Rumelhart, 1981; Rumelhart & McClelland, 1982; Balota & Rayner, 1991; McClelland, et. al., 1992].

It was decided to use the heterarchical model for the purposes of this thesis. This was because the existing sources of information would fit in with such a model (see section 1.8). However, a good alternative to the heterarchical model is the blackboard model. The blackboard model uses the metaphor of a blackboard [Reddy, 1975; McClelland, 1986]. Specialist knowledge sources write their results on a blackboard that can be viewed by all other relevant knowledge sources. Some knowledge sources can work from the bottom up, these are data driven knowledge sources. Other knowledge sources can work from the top down, suggesting hypotheses to account for the data that have arrived. These are conceptually based knowledge sources. Hypotheses may be wrong. A knowledge source may tell the system to expect things that never happen. The data driven knowledge sources must be able to correct the conceptually driven knowledge sources. All the knowledge sources have access to a centrally located blackboard. Each knowledge source monitors the blackboard for data that it can analyse. When a knowledge source recognizes data that it can analyse, it begins to process the data. Once a knowledge source has finished its specialized task, it writes the result on the blackboard for some other knowledge source to pick up. A general overall supervisor is also needed. The supervisor guides the specialist knowledge sources in a co-operative effort. An example of a script recognition system based on the blackboard model is that of Higgins & Bramall [Higgins & Bramall, 1993; Higgins & Bramall, 1994].

The handwritten address recognition system developed by Srihari and Keubert uses many different specialist modules: line separation, word separation, parsing, ZIP Code segmentation and recognition, postal directory access, word recognition and encode decision [Srihari & Keubert, 1997]. Word recognition combines a word shape recognizer which uses contour features with a character recognizer. The word shape recognizer is applied first. If the confidence obtained from the word shape recognizer is neither very high nor very low then the character recognizer is applied. If the character recognizer agrees with word shape recognizer then the result is chosen. The control structure also has an image reprocessing option. If a final choice is not reached then an image enhancement algorithm is applied and the recognition process begins anew.

A number of different approaches to the combination of knowledge sources can be observed. Perhaps the most straight forward of these approaches is to use a method based on voting. This can be based on a simple majority voting principle [e.g. Suen, et. al, 1990; Nadal & Suen, 1993], or on candidate subset combining and re-ranking [e.g. Ho, et. al., 1992b].

An alternative approach to the combination of methods is to draw upon knowledge of their individual characteristics. In this instance, some form of statistical or uncertainty reasoning can be employed.

Statistical approaches include a method based on Bayesian formalism [Xu, et. al., 1992]. Combining the votes of individual recognizers can itself be seen as a recognition task. Franke & Mandler treat recognition results as features and input them into a polynomial classifier [Franke & Mandler, 1992].

Uncertainty reasoning has relied upon the Dempster-Shafer theory of evidence [Mandler & Schurmann, 1988; Xu, et. al., 1992; Franke & Mandler, 1992]. Each individual classifier (or knowledge source) is transformed into a confidence value, which is used as the basic probability. Dempster-Shafer theory is used to combine the contribution of each individual classifier to give a final result.

Voting might appear to be a less sophisticated approach to the combination of methods than either statistical or uncertainty reasoning. However, Franke & Mandler report that only marginal differences exist between the results obtained using the Dempster-Shafer theory of evidence and combining votes using a statistical approach [Franke & Mandler, 1992]. Xu found that an approach based on Beyesian formalism proved unreliable [Xu, et. al., 1992]. In contrast, approaches based on Dempster-Shafer formalism and voting both behaved well. The Dempster-Shafer based approach was marginally than the voting method, especially when high reliability was required, but the difference in performance was slight.

2.4 Justification for the Word Level Cues used by the Word Level Method

2.4.1 Integration

The word level is a good area to begin the process of integration standing as it does between the letter level and the meta-word level. The word level is an intermediary between the pattern recognition side of script recognition and the language side. The word level can be seen to partake of both of these two different kinds of information. The word level utilizes information which is derived from pattern recognition (e.g. the exploitation of cues) and information which has a broader contextual foundation (e.g. lexical information and word frequency information). The word level can be considered the first point within the reading process that contextual information is applied. A movement from the contextual considerations of the word level to the broader contextual constraints of syntax and semantics is both a reasonable move and methodologically sound.

It is difficult to integrate these different sources of information because the various sources of information are so dissimilar and the kind of information which can be derived from them is qualitatively different. For example, semantic analysis takes a very different approach to handwriting recognition from that of the pattern recognizer. The information derived from semantic analysis is unrelated to the information derived from pattern recognition and therefore qualitatively different sources of information have to be combined in order to produce an integrated machine system. The mechanisms which have been used in this work for integration have been determined empirically. In this respect, the way in which the different sources of information are combined is arbitrary.

The various aspects of the reading process interact and any attempt to integrate these levels must take account of this interaction. It is precisely those areas where the different levels of the reading process overlap which are the special concern of anyone investigating the ways in which the various levels of information can be integrated in an effective and efficient manner. Visual cues and structural considerations (e.g. lexical constraints and orthographic regularity) influence the process of recognition in a bottom-up manner, whilst the addition of contextual cues increases the opportunity for top-down constraints to be exerted [Reddy, 1975; McClelland, 1986; Srihari & Bozinovic, 1987; Bozinovic & Srihari, 1989; McClelland, et. al., 1992; Higgins & Bramall, 1994].

Two different approaches towards the recognition of cursive handwriting can be identified. In the first approach, a pattern recognizer drives the recognition process. The main aim of this approach is the identification of what has been written. This aim means that the system is evaluated on the basis of whether or not it has produced the target word. Evaluation is primarily on the basis of one word output, and it is only of secondary importance that the system may give a list of alternatives in which the target word appears. Contextual information is not used. This is a bottom-up approach.

In contrast, the second approach uses the lexicon and applies contextual cues to select from this list of words. The method of evaluation in this approach is therefore whether the pattern recognition system gave the target word as an alternative and subsequently whether or not contextual cues make it possible to select the target word. This is a top-down approach. The reason why this second method is desirable is that some words are not well written and, in such cases, the second approach will find the word whilst the first method will not.

2.4.2 Selection of Cues for the Word Level Method

The following section explains how the human reading models described above lead to the word level cues used by the word level method. A number of factors which affect human recognition at the word level have been described (see section 2.2.4). It was decided that orthographic regularity, phonological regularity and morphological information would not be used in the word level method. The evidence for the use of morphemic information is weak (see section 2.2.7). Orthographic and phonological effects may be post-lexical. Given that this is the case, the information which orthographic and phonological regularity can provide is already supplied by the lexicon.

Nine sources of information have been selected for the word level method (see Table 2-1). Two of these sources of information can be considered to be higher level contextual sources: lexical and word frequency information. These two factors make a large difference to human performance but are not apparent from the physical information derived from pattern recognition (see section 2.2.5 and section 2.2.6, respectively).

Higher level contextual sources of information
1) lexical information
2) word frequency
Word level cues
1) word length in characters
2) first letter
3) last letter
4) presence or absence of ascenders
5) presence or absence of descenders
6) presence or absence of i-dots and j-dots
7) presence or absence of t-crosses and f-crosses

Table 2-1: Sources of information at the word level

Seven word level cues have been selected. All of these cues have been used, in one form or another, in previous recognition systems (see section 2.3.2 and section 2.2.8.2) [e.g. Earnest, 1962; Brown & Granapathy, 1980; McClelland, 1986; Mozer, 1987; Ho, et. al, 1992a; Higgins & Bramall, 1994; Powalka, 1995; Madhvanath & Govindaraju, 1998]. The seven word level cues were selected because of the strong evidence that they are used by human readers. The selection of these cues was driven by what is desired but also what could reasonably be obtained from the pattern recognizer. The lexicon was also examined to see what cues could be used to partition the lexicon efficiently without simply replicating the information used by the pattern recognizer, i.e. what cues could be used to generate a reasonably short list of candidates given that the middle letters of a word were not to be used. All of these sources of information are effective at the word level. The extent to which they have an impact is solely dependent upon the accuracy of the source of information.

The evidence that human readers use word length (measured by the number of characters) as a support to recognition is overwhelming (see section 2.2.8.3). Word length was therefore selected as one of the cues for the word level method.

The evidence that first letter and last letter are cues used by human readers is compelling. First and last characters are the most significant characters to human recognition (see section 2.2.8.2). Human readers appear to code the position of letters as end-letters or internal [Humphreys, Evett & Quinlan, 1990; McClelland, 1986; Mozer, 1987]. Since human readers use information about end-letters, the cues first letter and last letter were selected for the word level method.

The final four cues selected for the word level method were the presence or absence of ascenders, descenders, i-dots and j-dots, and lastly t-crosses and f-crosses. The four cues are intended to be a set of abstract cues which preserve the sort of information retained across letter confusions [Bouma, 1971]. Visual cues are obviously used by human readers (see section 2.2.8). It is true that case alternation studies suggest that letter information is not visually based (see section 2.2.8.1). However, parafoveal orthographic priming experiments suggest that middle letters do have a priming effect (see section 2.2.8.1). The effect of middle letters is position independent [Humphreys, Evett & Quinlan, 1990].

The four cues presence or absence of ascenders, descenders, i-dots and j-dots, and t-crosses and f-crosses, have been selected because letter confusions tend to share common properties of shape with the intended letter. There are problems with using letter confusions in the case of cursive script. The reason for this is the complexity of segmentation and the variability of letter forms (called "allographs"). Recognition errors tend to preserve certain cues because of a consistency between letter confusions and the intended letter, e.g. the fact that tall letters tend to be confused with other tall letters [Bouma, 1971]. However, no judgement is made about the specific kind of letter confusion or confusions, e.g. whether it is 'cl' -> 'd', or 'h' -> 'd'. These cues are not entirely physical. For instance, they are position independent. The words "abandon" and "abdomen" would have the same abstract representation under this scheme.

A pragmatic argument, rather than a strong theoretical one, has been used to select this exact set of cues. There are other alternative cues which could have been chosen. For example,

projecting lines, e.g. the characters 't' and 'r' both have a projecting line, whilst the character 'o' does not.

curves, e.g. the character 'c' has a curve on the left, 'b' has a curve on the right, whilst 'o' has two curves, one on the left and one on the right.

holes, e.g. the characters 'o' and 'd' both have an enclosed hole within the character, whilst 'c' does not.

The seven word level cues are not physical, or at least not entirely physical, but rather are abstract [Humphreys, Evett & Quinlan, 1990]. These cues are not simply physical features. Rather their relevance to recognition is with respect to identity rather than form. They are cues which survive iconic memory in human readers. The two cues first and last letter are not items of physical information. The end letters are treated as abstract representations. These two cues are characters, i.e. identities not physical patterns. It is not just characters which are represented internally using an abstract coding. Length is expressed in number of characters. Word length in letters is not a physical characteristic of the word, like shape, or the ratio of height to width, but is an abstraction. There is no simple relation to the actual physical length of the pattern, but an abstract representation of length based on the number of characters identified in a word.

Whilst the WSE is significant it is also apparent that sources of information other than those which appear to contribute to the WSE have to be used in order to ensure a reasonably correct reading of cursive handwriting. It is not expected that a simulation of the WSE will produce 100% correct recognition of cursive script but that it will serve to boost reading performance to a similar extent that it does with human readers.

2.5 Conclusions

This chapter described research into human reading and cursive handwriting recognition. The main area of the author's interest has been outlined. Information concerning the use of word level cues by human readers has been provided. The evidence that human readers use such cues is overwhelming. It is also apparent that human readers gain considerable benefit from their exploitation. Studies of human vision, in particular research concerned with parafoveal vision, highlight the cues used by human readers. Human readers use only a fraction of the available cues, e.g. parafoveal vision is less clear than foveal vision but human readers can recognize words using parafoveal vision alone. A consideration of the WSE has helped to identify the factors which affect human recognition at the word level. These factors are lexical information, word frequency information, orthographic regularity, phonological regularity, morphological information and word shape information. Word shape information has been further identified as abstract letter encoding, first and last letter recognition, and word length recognition.

A review of research into the computer recognition of cursive script has been presented. This review has concentrated on word shape recognition because this approach is the one most relevant to this thesis. A number of systems which use word shape information have been described.

Following on from the reviews of human and machine reading, a justification has been made for a set of cues to be used by the word level method. Some factors which may affect human recognition at the word level are not used. This is because the evidence for their use is weak, or because the use of a lexicon makes their use redundant. The selection of the cues was also made on pragmatic grounds. Consideration had to be given as to what could be reasonably obtained from the pattern recognizer.

Chapter 3: Establishing the Need for Word Level Context

3.1 Introduction

This chapter presents evidence for the usefulness of word level cues and provides supporting evidence of the need for contextual information. The validity of these arguments has been assumed, but has not yet been proven. The experiment given here proves that there is a context effect at the word level which is of relevance to the machine recognition of cursive script.

3.2 Experiment 1: The Word Superiority Effect

3.2.1 Introduction

It was observed in both Chapters 1 and 2 that human readers use contextual information to aid their recognition of cursive script and it was argued that a machine system must also use contextual information if it is to be as competent as human readers. It has been demonstrated that meta-word contextual information is of use to the machine recognition of cursive script [e.g. Keenan, 1993; Rose, 1994]. However, it has not yet been demonstrated that word level contextual information is relevant to the machine recognition of cursive script. Evidence for the relevance of word level contextual information must demonstrate two things. Firstly, it must be shown that it is possible to exploit word level information in order to improve recognition performance. This will be shown by proving that human readers perceive letters more accurately when they appear in words than when they appear on their own. Secondly, it must be shown that a machine system would be more efficient if it could utilize word level contextual information.

Given ambiguity at the pattern level, it is important to establish what kind of level of performance should be expected from pattern recognition alone. One way to estimate upper limits on performance is by studying how well the best reading system, that of the human reader, performs. Whilst it is not the case that human and machine must exploit the same information in the same manner, where a human reader cannot recognize a character reliably, the machine should not be expected to do so. The patterns of error are also of interest. If the human reader confuses certain letters with certain others, then the machine should make similar errors. If the machine system is developed so that it performs very well on certain samples of script, this may well be at the expense of other samples. If it recognizes a particular character which is not a good example of that character, this may cause it to make unexpected errors on other examples of that character. If a character does not look much like the character it is intended to represent it should not be recognized as that character. Humans make understandable errors on characters. For example, letter confusions tend to produce errors in which the pattern of ascenders and descenders is preserved [Bouma, 1971]. This may well reflect the processes involved in recognition. Since humans can read many diverse kinds of script, even those previously unseen, it is reasonable that machine systems should behave in a manner which bears some resemblance to humans.

Human readers perceive letters more accurately when they appear in words than when they appear in other contexts. This is known as the Word Superiority

Effect (WSE). An experiment was undertaken to examine the WSE in human readers for a sample of cursive script, and to compare their performance with that of the machine system on the same sample. The WSE is well established using printed text. However, there has been little research on the WSE using cursive script. This experiment demonstrates that the effect also exists with handwritten text and with a similar degree of influence.

A comparison between human and machine will fulfil two objectives. Firstly, it will indicate the level and nature of performance we should be expecting from the machine system. The machine system should be at least comparable in its performance to human readers at the task of recognizing handwriting which the reader is unfamiliar with, since one objective behind developing the current machine system is a system which is capable of dealing with unconstrained cursive handwriting and a large lexicon. Secondly, it will examine whether the lexical information implemented in the system is providing the type and level of help that human readers benefit from. This experiment presents a comparison between machine and human performance on the same samples of handwriting in order to observe what benefit human readers gain over the machine by their exploitation of word level information. The pattern recognition system developed in the Nottingham Trent University is competent. Human readers must be exploiting word level cues if they are better at recognizing words than the pattern recognizer, and if their performance on letters in isolation is similar to that of the machine system. However, the machine system should be able to use word level cues in order to improve performance. Chapter 5 will show that word level information can indeed be used by a machine system.

The main method currently used for comparing different systems is to look at their relative performance on a standard database of handwriting. A few such databases currently exist [e.g. CEDAR database, Essex database] and a large scale international project aimed at the laboratory benchmarking of pattern recognition is also underway [Guyon, et. al., 1994]. Comparing performance with that of humans offers another approach to this problem, and one which also makes some attempt to establish upper limits on performance.

There are only a few articles which investigate the processing of handwritten words by human readers [e.g. Corcoran & Rouse, 1970; Ford & Banks, 1977; Manso de Zuniga, et. al., 1991]. It has been assumed in this thesis that the processes involved in reading handwritten and typed words are identical, i.e. the route taken during the processing of script is the same regardless of what style it is, whether it is cursive handwriting, handprinted or typeface. For reasons why this should be the case, and evidence that handwritten and printed words are not analysed by separate processes, see Manso de Zuniga, et. al.. This is not to say that handwriting and typeface are necessarily treated in exactly the same way. For instance, Manso de Zuniga et. al. suggest that "handwriting requires extra processes, to deal with segmentation and item variability". The authors call these extra processes "cursive normalization" but also go on to suggest that typewritten words may not escape cursive normalization.

It is important that handwritten and typed words are shown to be read using identical processes since this means that evidence obtained from studies of printed words are also applicable to handwritten script. It is therefore reasonable to use such evidence with regards to cursive handwriting. For example, Manso de Zuniga, et. al. have shown that certain effects that have been well established on typewritten words also exist with handwritten words, e.g. the fact that word repetition facilitates identification and the existence of the word frequency effect. Indeed the authors found that the effects of word frequency tended to be greater on handwritten words than on printed words [Manso de Zuniga, et. al., 1991].

3.2.2 Method

A sample of lower case, cursive script was collected. Legibility was decided by consensus and machine performance. It was necessary that the machine performed reasonably well to provide some data. It was also required that human readers could read the script, although some errors were required to avoid ceiling effects, i.e. effects caused by an upper limit. For example, it would not be possible to demonstrate the WSE if recognition performance of individual letters was perfect. The handwriting of different people was first examined and the writer whose writing was fairly successfully recognized by the recognition software and was judged generally legible by human judges was chosen. A single writer was required so that the test data was consistent. Several specimens of the test set of words were obtained from the chosen writer in order to improve clarity. The handwriting was also rated by human readers for its general legibility. The data was to be presented to human subjects to recognize, so as to compare performance with that of the machine system. The comparison was made with the existing pattern recognition software, that is to say, without the addition of any word level information, apart from lexical, to the recognition process. The pattern recognizer had previously been trained on 4 single word examples of the writers' handwriting.

Two related sets of data were used: whole words and letters taken from these words. 26 words, with each letter of the alphabet contained in at least one of the words, were used as test data (see Appendix A). This will be called the 26 word data set. Each of the 26 words were selected so that the significant letter was inside the letter string, i.e. neither at the beginning nor at the end of the word. One of the central problems in the recognition of cursive script is character segmentation. It has been argued above that first and last letter may play a special part in the human reading process. For instance, the problem of segmentation is easier in the case of the first and last characters of a word since such letters do not have both a preceding and succeeding letter. Mid letters were therefore used in this experiment.

The target words were balanced for word length and frequency. The words were chosen to be of medium frequency (between 100 and 44 in the Kucera and Francis word frequency count [Kucera & Francis, 1967]) and to be between 4 or 5 letters long. It has been observed above that word frequency can be shown to make a contribution to the WSE. The use of medium frequency words prevented ceiling effects caused by any word frequency bias on behalf of the subjects. Double letter combinations were avoided. However, in two cases, it proved impossible to select target words that matched the criteria of word length, word frequency and position of target letter whilst also avoiding double letter combinations. These two words were "fell" (target letter 'l') and "pass" (target letter 's').

A further 10 words beyond the original set of 26 were also selected. These 10 words were between 4 and 6 letters long (see Appendix A). These words were selected for two letter combinations within the letter string and from the 10 words a corresponding set of 10 two letter combinations were segmented out. This was for the following reason. Some examples of letters can be read as one or two letters, depending upon their segmentation. What was required was that the subjects gave their first impression of what the stimuli represented. If they knew that only single letters were to be presented, their responses would be biased to reporting only single letters. To overcome this problem, pairs of letters were segmented out from the additional words. Since some stimuli represented two letters, and subjects were expecting one or two letters, it was hoped that subjects would feel able to report two letters if that was their impression. It was also necessary to avoid the possibility that the subjects would become aware of the alphabetical character of the original set and use this information to aid their identification. The addition of words and two letter combinations to the original data set reduced this possibility. None of the subjects reported being aware of the alphabetical basis of the data set thus avoiding any possibly distortion to the results obtained.

The test data was written on electronic paper (a Wacom tablet and electronic pen). It was captured using the recognition software performing as a stroke (or ink) collection program. The program was run on an IBM 486 PC. The data was stored for later processing by the recognition software running on a Sun 10 workstation. The data used in this, and other, experiments was derived from the handwriting recognition system developed within the Nottingham Trent University [Powalka, et. al., 1993]. The handwriting recognition system works by matching letter and segmentation patterns to a pattern database. The information was converted into a bitmapped graphics format for later display. In this format each word, and subsequently each letter or two letter pair, was stored as a separate image.

The significant letters from the set of test words were segmented out by hand, since multiple segmentations of a word by the pattern recognizer are possible and the selection of one particular middle or end letter has not been automated. Information for how each letter should be segmented out was taken from the recognition software. This was done so that a direct comparison between the performance of the handwriting recognition software and the performance of human subjects could be made.

The two sets of data (letters and words) were presented to subjects in exactly the same way. The data was displayed on a computer screen with the writing being coloured black on a white background. A short introductory sequence of 5 letters or words was used in order to help familiarise the subjects with the nature of the experiment. This sequence was the same for each subject. The other letters or words were displayed in a random order which was different for each of the subjects.

It was explained to each subject that the data was real human writing and that it was in lower case. Furthermore, the subjects were told in the case of the words that they were real, common English words and, in the case of the letters, that one or two letters could be expected. The stimuli were displayed as a word or a letter at a time following a fixation point, and each image was displayed on screen for 300 milli-seconds. This length of time is at the higher range of normal durations of fixational pauses in reading. This is long enough for the stimuli to be seen clearly, but not too long, to avoid subjects changing their minds. Subjects' immediate, first impressions were required in order to observe their direct unconscious response to the visual image which had been presented on the screen. The subject was then asked to type in what he or she had seen with no constraint on time being imposed. Only one response to an image from each subject was allowed.

The 12 subjects, 1 female and 11 males, were unpaid volunteers from the Department of Computing: graduates, post-graduates or technicians. Half of the subjects were shown the letters first and the words later on, and half the other way around. A gap of a day between being shown the one set of images and other was used in both cases. These precautions were taken to minimise learning effects. Subjects were aged between 22-46 years, were competent readers, and had normal or corrected to normal vision. The stimuli subtended an angle of up to approximately 3 degrees and thus were presented to foveal vision.

The second part of the experiment was an analysis of the performance of the handwriting recognition system on the stimuli which had been presented to the human subjects. All the experimental words were in the vocabulary of the system. A 15,000 word lexicon had been created by taking the 15,000 more frequent words [Johansson, 1980] from the Lancaster-Oslo-Bergen (LOB) Corpus of British English [Keenan, 1993]. The lexicon contained every morphological variant of each of the experimental words (see below). If a morphological variant was not already present in the lexicon, it was added to the lexicon. In total the lexicon contained 15,463 words. This lexicon is used in all of experiments reported in this work. This lexicon will be called the common lexicon. The maximum possible number of words in the list of alternatives generated by the pattern recognizer was limited to 10. The initial main part of the experiment only considers the top ranked choice of the pattern recognizer. The aim of the experiment was to make a direct comparison between machine and human performance and subjects were only allowed to provide one immediate response, so only one choice was required from the pattern recognizer. However, machine performance improves when lower ranked alternatives are also taken into account, but the recognizer tends to place the target word near the top of its list of alternatives in those cases where it has identified the target word. The performance of the pattern recognizer when the top 10 candidates are taken into consideration was therefore used in order to see what scope existed for utilizing other sources of information to select from a longer list of alternatives.

Each word was firstly presented to the recognition software and its response recorded. Secondly, the way in which each word had been segmented by the software was examined and the significant letter from the particular word under consideration selected according to this segmentation data. This allowed the software to respond with a ranked list of letter choices. The segmented letters were the same as those presented to the human subjects.

3.2.3 Results

All of the tables show a comparison between the recognition performance of the human subjects and the machine system.

Table 3-1 shows percent correct recognition for the subjects and the recognition software (not including results for two letter stimuli or practise trials). The column results are, in order, correct recognition of the individual segmented letters, correct recognition of the words and, lastly, the case of correct recognition of the words, or the word not being recognized but the significant letter in the word correctly identified, i.e. "caver" being seen instead of "cover" where 'v' is the significant letter. Note that the machine results reflect the top ranked choice.

	letter	word	letter in word
human	75.0%	85.6%	93.9%
machine	65.4%	61.5%	73.1%

Table 3-1: Percent correct recognition by human and machine

These results demonstrate the WSE with cursive script. The human data were analysed using a one way analysis of variance to test the effects of type of letter context. The effects of this factor were highly significant ($df(2,33)$, $F=33.2$, $p<0.001$). The words were recognized significantly better than the letters alone, with the letters in words being superior.

The machine performed better on letters in words than on letters alone, but was slightly worse on correct words. However, the difference between letters and letters in words is much smaller than that for human subjects, and the difference between letters and words, although small, is in the opposite direction. The difference in performance by the machine between letters in words and letters alone is probably due to word look-up.

Table 3-2 makes a comparison between the poorest human readers and the recognition system. The table shows how the machine system compared to the poorest human reader in the case of individual letters (this subject alone did not do as well as the machine), and also looks at the case of the poorest human reader for the full words.

	letter	word	letter in word
worst subject on letter recognition	61.5%	84.6%	96.2%
worst subject on word recognition	69.2%	73.1%	92.3%
machine	65.4%	61.5%	73.1%

Table 3-2: Percent correct recognition for the worst subject on letter recognition and the worst subject on word recognition vs. the machine

There were 15 words which every human subject correctly identified. The recognition software correctly identified 53.3% of the target words correctly identified by every human subject. The recognition software correctly identified a total of 16 words. The human subjects correctly identified 85% of the target words correctly identified by the machine.

3.2.4 Discussion

The results confirm the presence of the WSE for cursive script, since they show that letters in isolation, i.e. without any surrounding lexical, semantic or syntactic information, are not recognized as well as letters in the context of words by human readers (see section 3.2.1).

The most significant result was that people do not recognize individual letters perfectly. The machine system is not lagging that far behind human recognition of letters in isolation (65.4% for the machine, 75% for the human subjects). Indeed one subject did not recognize the letters as well as the machine did (65.4% for the machine, 61.5% for the subject).

However, the great advantage which the human readers had over the machine system was their ability to exploit cues at the word level. The machine system has access to information at the word level, but is clearly not using it in the same way as the human readers.

In part, the difference between the performance of the recognition software and that of the human readers can be explained, firstly, as a result of the software's inability to fully exploit diacritical and zoning information. Currently the system uses the presence of dots (e.g. above an 'i') to confirm the existence of a particular letter rather than to suggest its presence. Secondly, the software does not use all of the possible stroke information, e.g. the letter 'z' was crossed in the present sample and this appeared to help human readers decide that the letter was indeed a 'z' rather than, say, an 'l'. These problems with the software are being addressed. For instance, zone information is used, but cannot be extracted entirely accurately [Powalka, et. al., 1993].

The slight improvement shown by the machine at the word level is the result of lexical constraints. However, their effect is nowhere near that for human readers. It is also apparent, therefore, that such constraints are not enough by themselves to replicate the WSE in the machine system. It could be argued that the letter recognition performance of the machine is within the normal range, although towards the low end of it. There is obviously room for improvement when applying contextual constraints. The use of physical information can be improved, as noted above.

In the present machine system, word candidates are ordered on the basis of their physical characteristics. Performance is better when the top 10 candidates are taken into account, and a further slight increase in performance is reflected with even longer candidate lists. Table 3-3 compares percent correct recognition for the subjects, the recognition software when only the top ranked choice is considered, and the recognition software when the top 10 alternatives are also taken into consideration. When the top 10 alternatives are taken into account then the performance of the machine system is slightly above that of the subjects for whole words (88.5% for the machine, 85.6% for the human subjects), and only slightly below for letters in words (92.3% for the machine, 93.9% for the subjects). Whilst like is not being compared with like, these results suggest that the ordering of candidates is important. These results appear to indicate that if the machine could utilize word level cues then it would be performing at the same level of competence as human readers and would be displaying an effect akin to, if not necessarily the same as, the WSE for human readers.

	word	letter in word
human	85.6%	93.9%
machine (top choice)	61.5%	73.1%
machine (top 10 alternatives)	88.5%	92.3%

Table 3-3: Percent correct recognition by human, machine (top choice) and machine (top 10 alternatives)

This chapter has only considered the recognition of isolated words. As can be seen from Table 3-1, human subjects did not perform at the 100% level for whole words, even though there was a substantial WSE. While human handwriting can never be completely unambiguous, so that 100% performance cannot be expected, further improvement in performance can be gained by using context beyond the word level. Constraints of syntax and semantics can be used to aid recognition. It is information of this nature which enables disambiguation of words such Figure 1-7. The integration of meta-word contextual information is discussed in Chapter 6.

As far as expected levels of performance are concerned, human performance levels give a rough guide to what might reasonably be expected. There is variation between subjects. For letters, human readers are between 61.5% and 84.6% correct. For words, subjects range from 73.1% correct to 96.2% correct, and for letters in words from 84.6% correct to 100% correct. Average performance gives an approximate guide to what should be expected from the machine, with the best performance as a target. No subject got the letters 100% correct without the help of word level context. No subject recognized all of the words correctly. The differences between conditions occurred for all subjects. This indicates relative performance levels which should be manifest.

Performance at the pattern recognition level is rarely entirely certain; word level context can be used to bolster confidence. This will be particularly beneficial when confidence at the pattern recognition level is either low or undecided.

3.3 Conclusions

The experiment presented in this chapter showed that there is a context effect at the word level and that this effect is relevant to the machine recognition of cursive script. The experiment compared the performance of human readers and a machine system. Human readers displayed a much greater effect of word context than did the machine system. It may be the case that the use of physical information by the pattern recognizer can be improved. However, a comparison between human and machine performance on the same set of letters and words suggests that the performance of the machine system can be improved if it can utilize word level cues. Obviously there are differences between machine and human recognition of handwriting. The performance of the machine system when the top 10 candidates are taken into account indicates the expected improvement. It is therefore now necessary to turn to methods for deriving and exploiting such word level information.

Chapter 4: Developing the Word Level Method

4.1 Introduction

Human readers use word level cues when reading. Human readers often engage in only a limited examination of the written word. Human readers do not have to recognize individual letters perfectly. Human readers can use word level cues to derive a general impression of the word, and in some cases to identify it without more detailed examination [Rayner, et. al. 1982, O'Regan, 1979]. Higher-level context can be used in conjunction with word level cues to make assumptions about the word and from this to derive a candidate. More detailed examination of the word can be used to verify the choice. It has proven possible to indicate what sort of performance should be expected from pattern recognition software. This helps to identify the point at which further changes to any recognizer are futile and to identify the potential improvement given by the use of various contextual sources of information.

Experiment 1 has indicated that word level cues will help the machine system to reach a level of performance comparable to that of human readers. A number of cues which are useful at the word level have been set out. Previous work has shown that these cues can be employed effectively by a machine system [Bellaby & Evett, 1994b; Evett & Bellaby, 1994a]. One possible explanation for the WSE is that human readers limit the size of the lexicon which they draw upon. A small lexicon can be successfully combined with word level cues in order to identify a word uniquely. For example, it is possible to successfully identify a word using only its first letter in combination with an estimation of the length of the word. 32 words out of the common lexicon can be so identified. These include words such as "a", "do", "go", "no", "so" and "we", as well as longer, less common words such as "justification", and "worthlessness".

The need for word level cues has already been established. The method of applying the word level cues will be called the word level method. When known, word level cues are very effective in selecting words from a lexicon (see below). For a recognition system, however, the cues must be calculated from the information available, and will therefore not be 100% reliable because of variability, ambiguity and noise. The present chapter examines one way in which word level cues can be derived. The list of alternatives generated by the recognizer is examined and probable values for the cues calculated. The lexicon is then searched using these values. Viable candidates can be derived, even when the recognizer did not identify the target word. The present experiment investigated the influence of imperfect information about these cues on the performance of a cursive script recognition system. It also investigated a simple method for integrating these sources of information into a script recognition system.

A number of methods have been developed which use the output from a recognition system and try to correct it. Such methods effectively acknowledge that output from a recognition system may be incorrect but still contain useful information. Methods which attempt to identify and utilize this information in some manner include n-gram techniques [e.g. Riseman & Ehrich, 1971]; the application of letter and word context, such as the use of statistical information about letter sequences to calculate the most likely input word combined with a fixed dictionary [e.g. Ford & Higgins, 1990]; the Viterbi algorithm for producing the required output by removing or reducing unlikely letter sequences [Viterbi, 1967] and variants which improve upon this algorithm [e.g. Srihari, et. al., 1983]; string correction algorithms [e.g. Srihari & Bozinovic, 1982]; and methods grounded upon an acceptance of incomplete information (e.g. fuzzy sets and the use of wild cards). These methods can be seen to take one of two approaches. Either they are designed to correct single misidentified words, or they are intended for use with data preceding a dictionary lookup stage with the aim of avoiding such misidentification. These approaches are not appropriate to the kind of data extraction under discussion here, since the data is to be extracted from more than one word.

Another difficulty apparent with these approaches is that any attempt to deal explicitly with character confusions at the word level faces some formidable problems. For human readers it appears to be the case with type face that letter confusions tend to produce word errors in which the pattern of ascenders and descenders is preserved [Bouma, 1971]. However, it is not so clear whether this is also the case with cursive script. It is feasible to compile a list of probable word alternatives using common letter confusions and a lexical filter when the nature of the input text means that a single character will cause a single character confusion.

One major problem in the recognition of cursive handwriting is segmentation. Distinguishing a character from its preceding or succeeding characters in well formed, clearly spaced, typeface or handprinted script text is relatively easy. Since segmentation is not a particular problem, when character misidentification occurs one character will tend to cause a single letter confusion. This is not the case with cursive script. In the case of cursive script one letter may generate a single letter confusion, but it is also possible that double, or even triple letter confusions may arise, e.g. 'd' -> 'cl', or 'w' -> 'ill'. Contraction can also occur. Two characters can be misidentified as a single character, e.g. 'cl' -> 'd'. The presence of ligatures in cursive script also decreases the probability of a one to one relationship between the intended characters and the characters perceived in a misidentified word. The presence of ligatures does not just complicate the identification of single characters but also causes spurious characters to be perceived. Such spurious character generation is not easily predicted (if predictable at all), nor is it readily identifiable.

It is difficult to deal with these problems explicitly at the word level. Such an approach would mean the reproduction of work already done by the pattern recognizer itself. It would be, in that sense, redundant. It would also mean taking on board the segmentation problem faced by the pattern recognizer. Since double character misidentification and spurious character generation are possible, attempts to introduce confusions at the word level can become very complex, very quickly. It is only if the nature of the input text has the consequence that single characters cause single letter confusions, that it is feasible to compile a list of probable word alternatives using common letter confusions and a lexical filter. This is not the case, however, with cursive script. If allowance is made for single letters causing double letter confusions, and for two characters to be contracted down to a single letter, then the number of possible word alternatives increases dramatically. In other words, approaches at the word level which attempt to model confusability explicitly face the problem of combinatorial explosion. This problem is exacerbated by the possibility of spurious character generation.

The word level method (which is the method being applied here) provides a third, different, approach to the problem of recognition errors. For example, no attempt is made to correct an output word, instead new alternatives are generated. This method abstracts from the data generated by the recognizer. It is argued that this approach will be more effective than the other methods cited. The reason for this is that these methods attempt to rectify or avoid errors with single words, not a list of word alternatives. If the recognition process has generated a substantial error then it is, firstly, difficult to detect what kind of error has occurred and, secondly, very difficult to rectify the error, i.e. to reconstruct the intended word. In contrast, the word level method abstracts information from several words and therefore gives a better indication of what word was intended by the writer. Furthermore, one consequence of using a list of alternatives from a recognizer is that some of the problems of segmentation and character confusions are treated implicitly rather than explicitly, since the pattern recognition system (the 'expert' in this case) has already done most of the work involved. The list of words provided by the recognizer is the result of the confusions which the recognizer has identified and the alternative segmentations of the input which it has applied.

A more discriminatory way to include probable cue confusions and thus likely word candidates is to follow the type of approach used by human readers. In this manner, it is possible to devise a more directed and cogent method which utilizes information about human reading methods in conjunction with, firstly, knowledge of the confusions generated by the recognizer and, secondly, knowledge of the words which are present in a lexicon.

4.2 Lexical Selection

Word level cues will be used to derive a new list of alternatives to add to the existing ones generated by the pattern recognizer. Leaving aside word frequency information for the moment, it is possible to show how useful this kind of information can be within the recognition process and to indicate how they can be combined to good effect at the word level. The use of word level cues makes it viable to select new candidate words from the lexicon to add to the original candidate list. This is obviously of particular relevance to those situations in which the recognition software completely failed to give the target word as an alternative.

Seven cues are used in combination with the lexicon: word length, first letter, last letter and the presence or absence of ascenders, descenders, i-dots and j-dots, and t-crosses and f-crosses. These seven cues are surprisingly effective. Perfect detection of the seven cues alone can lead to the identification of a single word even in a relatively large lexicon. For example, just over 20 percent of the words in the common lexicon can be uniquely identified using the criteria of these seven cues and over 50% of the words are in groups which have 4 members or less.

Even with the less accurate figures which have been derived from an examination of the list of alternatives suggested it is rare to produce a list which is unmanageably long.

The common lexicon contains a significant number of low frequency words. If the size of the search space is ordered on the basis of word frequency then the proportion of words which can be so identified noticeably increases, e.g. very high frequency words such as "the".

Appendix B contains tables showing how the lexicon is partitioned by each of the seven cues.

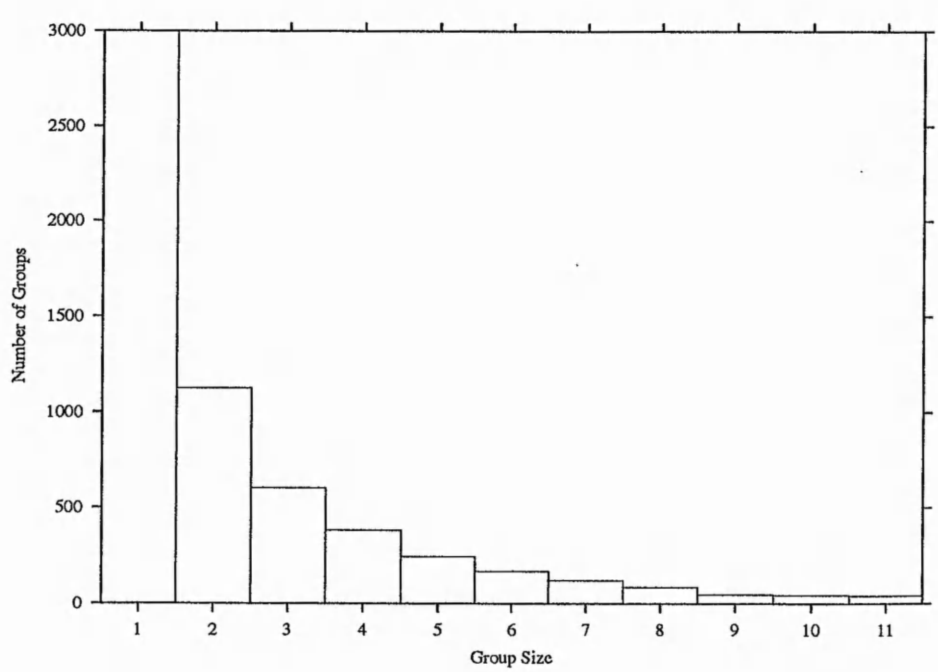
If all of the cues are brought together then a number of different groups of words which share the same pattern can be identified. The size of the groups identified by the seven cues vary.

The largest group of words delineated by the cues has 35 members, and only one group exists at this size. This group contains those words in the lexicon which have a length of 8 characters, which begin with letter 's', end with the letter 'g', and in which one or more ascender, descender, dot and cross are present, e.g. "settling", "shooting", "shouting", "staffing", "stamping".

Some groups only contain one member. This includes words such as "age" (length 3, first letter 'a', last letter 'e', descender present, but ascender, dot and cross absent), "able" (length 4, first letter 'a', last letter 'e', ascender present, but descender, dot and cross absent) and "abler" (length 5, first letter 'a', last letter 'r', ascender present, but descender, dot and cross absent).

The average number of words selected using the cues is 2.59.

Figure 4-1 shows the size of the groups identified by the seven cues, against the number of groups identified by the seven cues. In Figure 4-1 this data has been depicted using two histograms of different scale because the number of groups selected using the cues is substantially different for small group size and large group size.



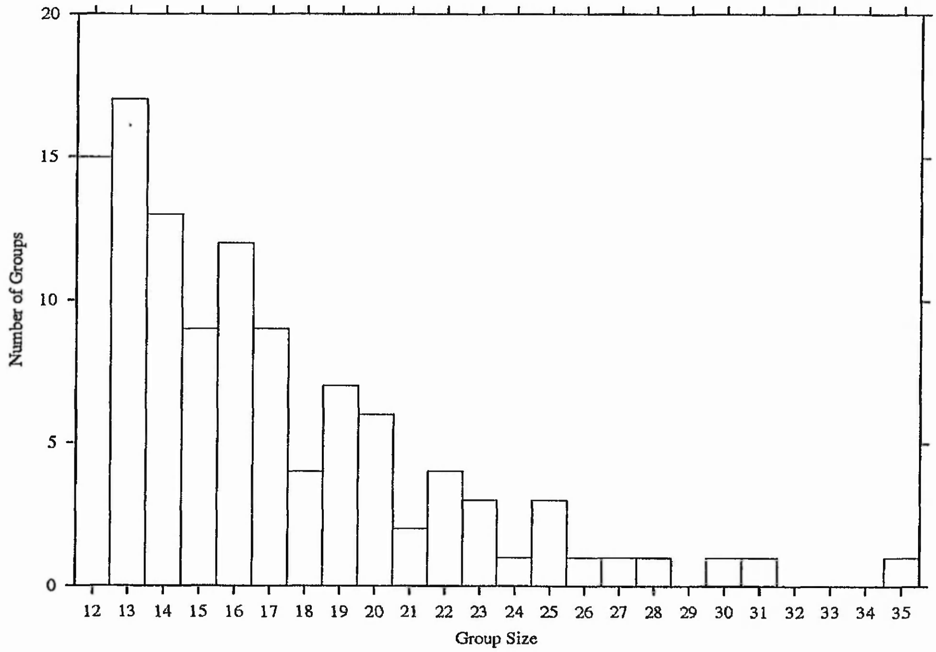


Figure 4-1: Size and number of groups identified by the seven cues

Assuming a maximum word length of 21 characters (the longest word length in the common lexicon) and using only the 26 characters of the alphabet, then the number of patterns has an upper bound of 227136 (21 x 26 x 26 x 2 x 2 x 2 x 2). However, about half of these patterns are invalid, e.g. it is not possible to have a word which has length of 1 character and whose first and last letter are different, or a word which begins or ends with the letter 'y' but does not contain the presence of a descender, since the letter 'y' is always considered to have a descender. In total, therefore, there are 116,058 valid patterns. However, only 5,966 of these valid patterns are used in the common lexicon. Unused words, and hence unused patterns, include pseudo-words such as "byd", "cib" and "molk".

4.3 Deriving the Cues

4.3.1 Introduction

A method for deriving word level cues from the list of candidate words generated by the pattern recognizer has been developed. Whilst it is possible to use pattern recognition methods to obtain values for these cues, this necessitates the development of a new system designed to carry out this task (see section 5.3). In this chapter no attempt to directly extract the particular cues under consideration from their input pattern is used. Rather, information about these cues is derived from the list of word alternatives given by the existing software. The reason it is possible to derive information about these cues from the candidate lists is that the pattern recognizer has already done much of the work involved in determining values for these cues. Take, for example, the number of characters in a word (the length). There are only so many letter confusions which can be made, even when expansion (one character confused with two or more) and contraction (two or more characters confused with one) are taken into consideration. Likewise, there are only so many ways in which a word can be segmented which results in real letters and valid letter strings. Whilst identification errors may mean that the top ranked choice is incorrect or that the target word is not one of the candidates in the list of alternative, the average word length of the candidates will tend to be similar to the target. A similar line of reasoning for all the other cues can be used.

The reason why valid information can be extracted from the list of alternatives is that the pattern recognition system has, within the constraints of the lexicon and errors caused by character misidentification and incorrect segmentation, attempted to produce the best set of matches to the intended word. It can be seen that the word candidates are related to the intended word. The kinds of error which the recognizer has made are not arbitrary, nor are they intractable.

A cursory examination of the data (e.g. consideration of the top ranked words only) might suggest that the output bears little resemblance to the intended word. For the pattern recognition system developed in the Nottingham Trent University this lack of resemblance is a consequence of substitution errors caused by misidentification of characters and exacerbated by the lexical lookup routines.

The Nottingham Trent pattern recognizer may have identified individual cues of the intended word with a high degree of accuracy. However, candidate selection is dependent on the way in which these cues are combined. A characteristic of those cases where the recognition software has completely failed to give the target word as an alternative is that the cues are incorrectly combined. In consequence, each individual word in the list bears little resemblance to the target word. Every word in the list will contain one or more cues which are wrong and the extent to which the cues are wrong in each of the word alternatives is significant.

The reason why the intended word has not been produced by the recognizer is that the recognizer attempts to identify each and every character of the word. Misidentification of any one letter will necessarily mean that an error is produced. The output of the recognizer is restricted by the character confusions the recognizer has created.

The set of factors which generate this kind of error are as follows. The recognizer misidentifies one or more characters. Even without misidentification a substitution set will exist. Output from the recognizer will therefore contain character strings which are partially in error. Since the recognition process includes a dictionary lookup stage, this means that there will be whole-word substitution errors. There may be no strong resemblance between any one word in the list of alternatives and the intended word. However, some semblance to the intended word will probably exist. This may not be immediately apparent since it may be, for example, that only word length and the pattern of ascenders and descenders has been preserved. However, if more than one word is examined then it can be seen that a meaningful proportion of the shape of the intended word has been retained.

Although in some cases the recognizer generates very poor output, an important finding is that the recognizer tends to produce errors in which the shape of the intended word has been preserved. Reconstruction of the intended word should be possible when several words from the list of alternatives are considered together. Resemblance to the intended word tends to decrease further down the list, i.e. towards those words which the recognition system has given a lower confidence score. Errors which are primarily the result of partial word identification and lexical lookup are rectifiable, whilst those which are a result of a significant, or complete, failure to identify the intended characters are not.

In those cases where the target word has been included in the list of alternatives it is obvious that information about the shape of the intended word is present in the list of alternatives. This is not so obvious in those cases where the recognizer has completely failed to identify the target word. A preliminary examination of the word lists appeared to show that the words they contained bore a resemblance to the target word. For example, it was often possible to see that at least one word in the list of alternatives began with the correct letter. Likewise, it was apparent that the length of the words were approximately correct. What was not immediately apparent was whether the kind of information which could be perceived in one list had the same, or similar, characteristics to the kind of information apparent in other lists. One general method which could be used successfully to extract the desired information over a wide variety of different word lists was required. A variety of different methods for extracting the relevant information were devised.

Seven cues are used by the word level method

word length in number of character

first letter

last letter

presence or absence of ascenders

presence or absence of descenders

presence or absence of i-dots and j-dots

presence or absence of t-crosses and f-crosses

There are a number of different methods, variants, and thresholds which have been used to derive values for the cues from the candidate list.

deriving values (see section 4.3.2)

using the arithmetic mean

using the median

using the mode

forcing a single-valued outcome (see section 4.3.3) *by rounding fractions*

by reducing the list of alternatives

by choosing the value which appears highest in the list of words

limiting the number of candidates (see section 4.3.4)

using the ranked position of the candidates

using the confidence score of the candidates

weighting candidates by confidence score (see section 4.3.5)

Descriptions of these different methods are given below.

4.3.2 Deriving Values

The list of alternatives generated by the recognizer is examined and cue information for each of the words is extracted. Three different extraction methods were tested:

the arithmetic mean

the median

the mode

The mean, median or mode can be used to calculate an average value for the cue or to select its most common occurrence in the list of word alternatives. The cues first letter and last letter can only be derived using mode because the mode is the only measure of central tendency that can be used with unordered qualitative variables such as letters. Word length, ascender presence, descender presence, dot presence and cross presence can be calculated using any of the three selection methods.

For example, consider the list of word alternatives given in Table 4-1.

candidate	ascender	descender	dot	cross	word length	first letter	last letter
fire	1	1	1	1	4	f	e
lie	1	0	1	0	3	l	e
tie	1	0	1	1	3	t	e
fill	1	1	1	1	4	f	l
fit	1	1	1	1	3	f	t
tire	1	0	1	1	4	t	e

Table 4-1: Example list of candidate words

In the case of ascenders, descenders, i-dots and j-dots, and t-crosses and f-crosses a word receives a 1 if the property is present (however many times) and a 0 if it is absent, e.g. the word "fill" is given a score of 1 for presence or absence of ascenders even though it contains three ascenders (an alternative method is to use a count, see below). An instance of a category in the candidate list increases the observed frequency of that category by one. Similarly if the result of calculating the mean, median or mode is 1, this indicates that the property is present, whilst a result of 0 indicates that it is absent. Results for each of the extraction methods on each of the cues are given in Table 4-2.

	ascender	descender	dot	cross	word length	first letter	last letter
mean	1	0.5	1	0.83	3.5	-	-
median	1	0.5	1	1	3.5	-	-
mode	1	0 or 1	1	1	3 or 4	f	e

Table 4-2: Results for the example list of words

The arithmetic mean is the sum of the scores divided by the number of scores. The mean for the cue presence or absence of ascenders in the example is therefore

$$(1+1+1+1+1+1)/6=1$$

The mean for the cue presence or absence of descenders is

$$(1+0+0+1+1+0)/6=0.5$$

The median is the point in a distribution that divides the data in two groups having equal frequency. If the number of scores is odd, then the median is the middle score when scores have been arranged in order of size. If the number of scores is even, then the median is the midway point between the two middle scores. In case of the cue absence or presence of crosses it can be seen that 0 occurs once, whilst 1 occurs five times. When the scores are ordered from smallest to largest it can be seen that median is midway between 1 and 1

0 1 1 1 1 1

The median for the cue absence or presence of crosses is therefore 1.

The median for the cue word length is a decimal. It can be seen that 3 and 4 both occur three times. Arranging the scores from smallest to lowest gives

3 3 3 4 4 4

The median is therefore midway between 3 and 4, i.e. a result of 3.5.

The mode is the score that occurs with the greatest frequency. The mode for the cue last letter is calculated as follows. It can be seen that 'l' occurs once, 't' occurs once, and 'e' occurs four times. The mode is 't', since 't' occurs with the greatest frequency. Some distributions cannot be described by a single mode since two or more scores can have the same maximum frequency. The mode is therefore unable to measure the central tendency. For instance, in the case of the cue absence or presence of descenders it can be seen that 0 and 1 both occur three times. The mode is therefore 0 and 1.

An alternative to ascribing a score of 1 to a word if ascenders, descenders, i-dots and j-dots, or t-crosses and f-crosses are present is to count the actual number of times that a particular cue is present in a word and to use this score when determining the central tendency. If the resulting average is 0, or smaller than 0.5, then it is determined that the cue is absent, otherwise it is determined that the cue is present. For instance, in the above example the number of ascenders in the words is 1, 1, 1, 3, 2, and 1. The mean for the cue presence or absence of ascenders is therefore 1.5 which indicates that an ascender is present. This approach is obviously not appropriate for the mode, but it can be used for determining the mean and the median. Counting the actual number of times a cue occurs in a word is not as accurate an indicator as simply recording whether or not the cue is present. This is because it tends to make the presence of a cue more likely than is actually the case.

4.3.3 Forcing a Single-valued Outcome

It can be seen from the example given above that the mean and median can produce decimal results. It is possible to use decimal results, e.g. the input pattern could contain part of a letter. However, human readers encode letter identities, not their visual form. The pattern of cues used in the word level method correspond to an abstract representation of a word, not to the physical form of a word. For the purpose of this thesis it was decided, therefore, that only integer outcomes were required, e.g. it was not possible for a word to have a length (measured by the number of characters) of 3.5. It is highly likely that the mean will have a fractional outcome. The median can have a fractional outcome when there are an odd number of scores.

It is also possible for the mode to produce two or more outcomes with the same maximum frequency. In such cases the mode has been unable to measure the central tendency. Such a result is useless for the purpose of deriving values for the cues, e.g. if the mode for the presence or absence of descenders is both 0 and 1 this suggests that descenders are both absent and present.

There are three ways in which the extraction methods can be forced to generate a single outcome or an integer outcome.

1) Round fractions either up or down to the nearest integer value. This still leaves open those situations where the fraction is exactly .5. However, there is a convention in mathematics that when the digit to be dropped is 5, the digit to the left of the 5 is increased by 1 if it is odd, but left unchanged if it is even, and this convention was adopted in these cases [Kirk, 1990]. In the example given above the result of calculating the mean length of the words in the list was 3.5. Applying the mathematical convention, this decimal value would be rounded up

to 4. Only the mean is suited to this method of rounding fractions up or down since the existence of more than one value in the candidate list will always have the consequence that the mean is a fraction.

2) Reduce the list of alternatives by progressively removing its lowest ranked candidate until only one value is generated by the selection method. For example, the cue absence or presence of descenders in the example list of words produced a mode of 0 and 1. The lowest ranked candidate ("tire") is therefore removed from the list and the mode recalculated producing a single valued result of 1.

3) Choosing the value which appears highest in the list of words. Fractions are rounded both up to the nearest integer value and down to the nearest integer value and whichever one of these two values appears first in the candidate list is picked. The reason for this is that higher ranked candidates are more likely to be correct. In the case of the mode all of the values which have the same maximum frequency are recorded and the value which appears highest in the candidate list is selected. For example, the cue word length in the example list of words produced a mode of 3 or 4. The list of words is examined and since a word length of 4 appears higher than a word length of 3 in the candidate list (the top ranked candidate has a word length of 4) this is the value which is chosen. It is possible, in the case of the mean and the median, that the selected values do not appear in the candidate list, in which case the reduction procedure can be first employed until the method produces a value which does appear in the candidate list.

It is not possible to average the cues first character and last character. In this case it is only possible to reduce the list of alternatives, or to pick the highest appearance in the list of alternatives, in order to force a single valued outcome. In the case of the cues word length, ascender presence, descender presence, dot presence and cross presence the median and the mode can be forced to generate a single integer outcome using any of the above methods. It should be noted that strictly speaking it is statistically wrong to force the mode to generate a single outcome since two nonadjacent scores with same maximum frequency cannot be described by a single mode. Pragmatic reasons call for the mode to generate a single outcome rather than proper mathematical reasons.

4.3.4 Number of Candidates

The maximum possible number of words in the list of alternatives generated by the pattern recognizer was set to its maximum value of 100 alternatives. This was done so as to include instances where the pattern recognizer gave the target word as an alternative even though the target word was ranked low. This value of 100 may not be the best limit. It may be the case that reducing the number of candidates will produce results that are more accurate. The words generated by the pattern recognizer should bear some resemblance to the target word. However, it should also be the case that words ranked lower down the list of alternatives should bear less resemblance to the target than words ranked higher up the list.

Appendix C contains two figures showing the degree of resemblance between the candidates and their target word, firstly by ranked position of the candidate, and secondly by the confidence score of the candidate. The number of cues in each of the candidate words which were identical to the cues of the target word were recorded. A comparison between the number of cues which were correct and the ranked position of the candidate shows that the degree of resemblance does, on average, decrease the lower the candidate is ranked. For instance, there are no candidates ranked lower than 78 which have all seven cues identical to the target word. A similar picture also emerges in a comparison between the number of cues which are correct and the confidence score of the candidate. In both cases whilst resemblance does, on average, decrease further down, the decrease in resemblance is not as extreme as one might perhaps expect.

The pattern recognizer is attempting to recognize factors other than the seven cues concerned here, e.g. it is attempting to recognize all of the characters in the word (not just first and last). This in part accounts for the reduction in resemblance not being as steep as one would expect. However, it also indicates that the recognizer is not as sensitive to the relative worth of its proposed candidates and lacks a strong ability to identify those candidates which least resemble the target. Not surprisingly, there was no simple way to filter out candidates which completely failed to resemble the target word, or which only resembled it to a small degree. Any attempt to remove candidate words which did not resemble the target word using a simple filter also removed candidates which contained some reliable information about the target. It has not proved possible so far to determine whether such information is important and it seems unlikely that it would ever be possible because candidates from the pattern recognizer should bear some semblance to the target word even at lower ranks. It is not possible to determine *a priori* whether such information is important, e.g. it may duplicate information which is held further up the list.

The candidate list generated by the recognizer for each target varies in size. Each candidate is given a score which indicates the confidence of the recognizer in each of the candidates that it has generated. The list of candidates is ranked according to these confidence scores. The order of the list of candidates thus reflects the relative confidence of the recognizer that a particular candidate is the target word. It should be the case that candidates lower down the list of alternatives least resemble the target word, specifically those candidates which have the lowest confidence score should least resemble the target. However, it is not necessarily the case that a word's score and a word's ranking within the list of alternatives reflect exactly the same information since, for example, it is possible for every word in a list of alternatives to have a relatively low score. Since a lower score and lower rank should mean that the candidate word bears less resemblance to the target word it may be the case that ignoring candidates with a low score or with a low rank will improve the accuracy of the data derived from the candidate list.

Two different ways to limit the number of candidate words were considered.

1) a threshold based on the ranked position of the candidates;

2) a threshold based on the confidence score of the candidates.

The actual level of any threshold which is going to be the best for any particular method is not solely a function of the average degree of resemblance. The method also has to use enough of the candidate words in order to derive useful information. So there are two contrary impulses: increasing the number of candidates so that the maximum amount of information is obtained, and decreasing the number of candidates so that accurate information is obtained. It is at the point where these two different requirements meet that the best threshold will be obtained. The value of the thresholds used to limit the number of candidates will be determined empirically.

4.3.5 Weighting by Confidence Score

The influence of candidates with higher confidence scores when calculating the mean, median or mode can be increased. The candidates in the list of alternatives have an associated confidence score. It is possible to weight the mean, median and mode using these confidence scores. Words further down the list bear less resemblance to the target word. The score given by the recognizer to each word is meant to indicate how confident the recognizer is that the candidate resembles the target word. Weighting an instance of a cue according to a value derived from the confidence score of its source word should therefore be useful. For instance, the mode is the category or class that occurs with greatest frequency. As described in section 4.3.2 an instance of a category in the candidate list increases the observed frequency of that category by one. Weighting is implemented by increasing the observed frequency of an instance of a category in the candidate list by its confidence score.

For example, consider the example list of word alternatives given above in Table 4-1. Table 4-3 below reproduces the list of candidates with just their word length but also includes the confidence scores given to each candidate by the pattern recognizer.

candidate	word length	confidence score
fire	4	90
lie	3	70
tie	3	68
fill	4	59
fit	3	52
tire	4	48

Table 4-3: Example list of candidates including confidence score

The mode for the cue word length would be calculated using the method in section 4.3.2 as follows. It can be seen that 3 and 4 both occur three times. The mode is therefore 3 and 4.

When the confidence score is taken into account the mode is calculated as follows:

1) For each category

   1.1) For each instance of a category in the candidate list

      1.1.1) Increase the observed frequency of that category by the confidence score of the candidate

2) Determine the highest frequency score

3) Make the category, or categories, with the highest frequency score the mode

If the confidence score is taken into account then 3 occurs with a confidence score of 70, 68 and 52, which gives it a frequency score of 190 (70 + 68 + 52). 4 occurs with a confidence score of 90, 59 and 48, which gives it a frequency score of 197 (90 + 59 + 48). The highest frequency score is 197 and the mode is therefore 4.

If the recognizer is accurate at determining a cue, then weighting will increase the accuracy of the mean, median and mode. The reason for this is that candidates will have their confidence scores adjusted according to the determination of the cue by the pattern recognizer and therefore weighting should reflect this adjustment.

4.3.6 Experiment 2: Establishing Initial Parameters

4.3.6.1 Introduction

The list of alternatives generated by the recognizer is examined and cue information for each of the words is extracted. All of the different extraction methods were tested. The mean, median or mode can be used to calculate an average value for the cue or to select its most common occurrence. Words further down the list bear less resemblance to the target word, so various thresholds for the maximum number of words were tested. The score given by the recognizer to each word is meant to indicate how confident the recognizer is that the candidate resembles the target word. The usefulness of applying this information to the values produced by cue extraction was also tested.

4.3.6.2 Method

A sample of writing from 18 subjects was used, each of whom wrote down the same 200 words using lower case cursive handwriting (a total of 3,600 samples). The 18 subjects, 2 females and 16 males, were unpaid volunteers from the Department of Computing: post-graduates or lecturers. All of the subjects were competent adult writers with no writing difficulties. All of the writers were familiar with the task of writing on electronic paper using an electronic stylus. All the experimental words were in the vocabulary of the system. This will be called the 200 word data set (see Appendix D). The list of 200 words was designed to be representative of a large vocabulary. Relevant factors within the 200 word data set and within the common lexicon have a similar distribution. Comparisons between the 200 word data set and the common lexicon for letters of the alphabet and for all the word level cues are given in Appendix E.

It was required that the writing was neither extremely sloppy nor overly neat. The subjects were asked to write at their normal writing speed, using their normal writing style. The legibility of the data ranged from neat, but not overly neat, to poor, but not badly formed. Legibility was first determined by the writer. Each of the writers was given the opportunity to rewrite any words which he or she felt was untypical of his or her normal, clear handwriting. Secondly, legibility was determined by human judges with the aim of replacing any words which the judges felt would be illegible to a competent human reader. The judges were Robert Powalka and myself. The data also represented a range of writing styles.

The test data was written on electronic paper (a NCR 3125 pen computer and associated NCR electronic stylus). It was captured using the recognition software, i.e. the front end of the recognizer simply used as a data collector. The data was captured at a rate of 200 samples per second at a resolution of 100 points per millimetre. The recognition software encoded the writing in the form of stroke information and this data was stored for later processing by recognition software running on a Sun 10 workstation.

The pattern recognition software had already received a limited amount of training on other examples of handwriting from some of these 18 subjects.

The recognizer had been trained using between 80-90 words for 5 of the subjects, using 20 words for one subject, and using between 5-10 words for 3 of the subjects. The recognizer had received no training at all for the other 9 subjects. It was therefore possible to observe recognition results on a range of writers, from writers on whose writing the pattern recognizer had been trained to writers on whose writing the pattern recognizer had not been trained.

The performance of the pattern recognition software on all of the samples was recorded.

The pattern recognizer can, on occasion, completely fail to generate any candidates at all. These occasions will be called catastrophic failures. The pattern recognizer experienced a catastrophic failure in a total of 512 cases. The candidate list cannot be used to derive cues when the pattern recognizer has experienced a catastrophic failure. For the purpose of this chapter those occasions where the pattern recognizer completely failed to generate any output have been ignored.

The number of words in the list of alternatives was limited to a maximum of 100. The length of the candidate list ranged from 1 word alternative up to the maximum value of 100 word alternatives. The maximum possible number of candidates generated by the pattern recognizer was 100. The lowest rank at which a target word appears is 54. A limit of 100 candidates is therefore nearly twice as great as the lowest ranked target word. The legibility of the data ranged from neat to poor handwriting (target words top ranked by the pattern recognizer ranged from 92% for the best writer, down to 19% for the worst writer). The data also represented a range of writing styles. The common lexicon was used.

The reason for using a variety of different methods of calculation is to avoid a simplistic replication of the information held in the initial list. A range of different list sizes and selection methods were tested and evaluated to discover which combinations produced the best results empirically. A number of factors are involved in choosing the best combination of methods. The word shape depicted by the seven cues could describe more than one word. Cue detection was not perfect and therefore the introduction of confusions would extend the list of candidates generated by the method. The pattern recognition system had a reasonably low average error rate and an extremely low error rate on the writing of some of the subjects. It is significant that a fine degree of accuracy is not necessary. The reason for this is that in a great many cases even wide variance will only generate a low number of alternatives.

Other uncertainties exist in combining the cues. These uncertainties will affect the results, but it is difficult to delineate precisely their degree of influence at this stage. There are relationships between some of the cues. The set of words in the lexicon described using one cue can intersect the set of words described by a second cue. For example, in the word "a", first letter detection and last letter detection are the same; word length detection will also have a strong bearing upon such single character words. In a similar manner, the set of words described by the criterion "contains the letter 'd'", intersects the set described by the criterion "contains an ascender". There will also be interaction between cues. Since cue detection is not 100% accurate, there will be forced choices in those cases where one or more of the cues are contradictory, e.g. "contains the letter 'd'" and "does not contain an ascender".

4.3.6.3 Results

A variety of different methods and variants have been described. It is not practical to provide results for each and every one of these alternatives. For example, there are three ways to derive values, three ways to force a single-valued outcome, two ways to limit the number of candidates, and the calculations can be unweighted or weighted. The value of the threshold used to limit the number of candidates has a range of 100. Clearly, the number of different outcomes is large. Since there are such a large number of parameters, the results concentrate on limiting the number of candidates by their score or by their rank, and for the calculations being unweighted or weighted. These two parameters have been selected because they are related to the activity of the pattern recognizer in a way that the other methods are not. For example, the confidence scores are generated by the pattern recognizer, and the ranking of the word alternatives is dependent upon their confidence scores. More detailed results are given in Appendix F.

It was decided to use only part of the data sample for testing. Catastrophic failures and uninformative (see below) word lists were excluded, and only those lists in which the target did not appear, or was placed below rank 3, were used in the initial experiments. The number of lists used was 607. The total number of candidate words was 11,730, an average of 19.3 words per list. This will be called the partial data set.

The reasons for this were, firstly, that the pattern recognizer can experience catastrophic failures and therefore the candidate list cannot be used to derive cues. Those occasions where the pattern recognizer completely failed to generate any output were ignored (a total of 512 cases).

Secondly, the pattern recognizer produced a proportion of output which was wholly uninformative. Even though the recognizer had generated some output, the output was considered to be too inaccurate to be of any use. The use of such uninformative word lists in the word level method was inappropriate and would have lead, for instance, to the generation of irrelevant confusions and probabilities and therefore would have distorted the experiment. A small number of cases were therefore removed since their associated candidate lists were considered to be too inaccurate to be of any use (6 in total).

These six cases are shown in Table 4-4. Of these six cases, four candidate lists do not contain a single character from the target word, whilst the other two only contain one character from the target word. The candidate words also have a short length. These six candidate lists look like the type of output generated by the pattern recognizer when the target is a single character word. An example of this kind of output is given in Table 4-5. It was decided that the input data was so poor that the pattern recognizer had, in effect, treated it as a single character.

target word: power	target word: very	target word: good	target word: hundred	target word: important	target word: view
i	l	l	i	l	l
j	j	j	j		j
ti	y				f
	ill				
	if				

Table 4-4: The six uninformative word lists

target word: a
a
u
n
e
ill
d
if
o
it
f

Table 4-5: An example of output from the pattern recognizer when the target is a single character word

Thirdly, previous work had shown that the word level method tended to be most effective at recognizing, or improving the ranking, of target words which were ranked fourth or worse by the pattern recognizer. Therefore, it was decided to concentrate on improving the performance of cue derivation when the target word was ranked low or did not appear, rather than be concerned with the performance of cue derivation when the target word was ranked high.

The word level cues, and the word level method, should not simply follow in the footsteps of the pattern recognizer. The inclusion of instances where the target word was highly ranked would mean that cue derivation from the candidate list would tend merely to replicate, in a less discriminatory form, the output of the pattern recognizer rather than generate new information. For example, simply using the cues of the top ranked candidate to derive values for the cues would be an accurate method of calculating the cues because of the number of targets top ranked. This would mean that the target word would tend to appear in the list of word alternatives generated by the word level method. However, the word level method is not as discriminatory as the pattern recognizer since it selects groups of words from the lexicon using the criteria of the seven chosen cues. The word level method cannot hope to compete with the pattern recognizer in its ability to place the target word at, or near the top of, the ranked list of word alternatives because the word level method does not have the selectiveness of the pattern recognizer.

Firstly, a threshold was selected for each of the methods used to limit the number of candidates. It was decided to use the threshold which, on average, was the most accurate for all of the cues combined. Detailed results are given in Appendix F. Tables F-1, F-2, F-3 and F-4 show results, respectively, for rank without weighting, rank with weighting, score without weighting, and score with weighting. Results are for the partial data set. Results are given for the full range of the threshold used to limit the number of candidates. For all of the other methods, the combination of alternatives that produced the most accurate results for a given cue have been used. The value of the thresholds used to limit the number of candidates for the different combinations of methods are given in Table 4-6.

method	threshold
rank unweighted	8
score unweighted	49
rank weighted	7
score weighted	35

Table 4-6: Thresholds for the different combinations of methods

Secondly, the combination of methods for each cue was chosen. Appendix F gives the detailed results. Tables F-5, F-6, F-7 and F-8 show results, respectively, for rank without weighting, rank with weighting, score without weighting, and score with weighting. Results are for the partial data set. Results are only shown for the threshold used to limit the number of candidates. However, results for all of the other combination of methods are provided.

There was no strong indication that one method, variant, or threshold consistently favoured any particular cue.

When weighting was not used, limiting the number of the candidates by their score was, in general, more accurate than limiting the number of the candidates by their rank. This suggests that the confidence scores of the candidates are a more reliable indicator of their resemblance to the target than their rank.

When weighting was used, limiting the number of the candidates by their rank was, in general, more accurate than limiting the number of the candidates by their score.

Weighting improved the accuracy of cue detection for the rank method for all of the cues.

In contrast, the score method combined with weighting was less accurate than without weighting for six of the seven cues. Presumably limiting the number of candidates by their confidence score conflicts with the aim of increasing the influence of candidates with higher confidence scores.

The results for the chosen threshold using the combination of alternatives that produced the most accurate results are given in Table 4-7. The specific combination of alternatives for each cue will be given below. Results are for the complete 200 word data set, excluding catastrophic failures (a total of 3,088 cases).

method	length	ascender	descender	dot	cross	first	last
rank, unweighted	61.2%	90.1%	88.3%	88.4%	88.2%	69.9%	72.0%
score, unweighted	70.4%	89.2%	89.6%	89.8%	88.6%	70.4%	74.4%
rank, weighted	70.5%	91.1%	88.9%	89.4%	89.1%	71.9%	75.4%
score, weighted	69.7%	88.0%	87.9%	89.3%	88.7%	68.9%	73.3%

Table 4-7: Percent correct of the cues using the two methods of limiting the number of candidates- weighted and unweighted

The rank method combined with weighting was, in general, the most accurate combination of methods. The evaluation criteria described above was used to select the best combination of methods, variants and thresholds. The values were derived in the following ways:

word length: the median, reducing the list of alternatives until one most frequent value is left.

first letter: the mode, choosing the value which appears highest in the list of alternatives.

last letter: the mode, reducing the list of alternatives until one most frequent value remains.

ascenders: the mean, rounded down at 0.5.

descenders: the median, choosing the value which appears highest in the list of alternatives.

dots: the mode, choosing the value which appears highest in the list of alternatives.

crosses: the mean, rounded down at 0.5.

4.3.6.4 Discussion

A method for deriving word level cues from the list of candidate words has been described. There are many different methods which could be used to derive values for the cues. This experiment has demonstrated that information about word level cues can be successfully derived from the candidate list generated by the pattern recognizer. Cue detection is imperfect but it is accurate enough to be of use to the word level method. The accuracy of cue detection ranged from 91.1% for the detection of ascenders, down to 70.5% for the calculation of word length. Choices of parameters, values, thresholds, etc. have all been made empirically. Further work is necessary to establish optimum values. The main consideration at this stage was to demonstrate the effectiveness of the approach.

The confidence scores ascribed by the recognizer need to be more sensitive to the requirements of contextual analysis. There are two different kinds of sensitivity which need to be considered: firstly, the confidence score given to each candidate in relation to the confidence score given to every other candidate in the same list of alternatives; and secondly the relative confidence scores between different lists of alternatives. In those cases where the recognizer has identified the intended word, then it tends to place the word at the top of its list of alternatives. The recognizer is therefore responsive to visual cues in the context of a single word, i.e. it strongly displays the first kind of sensitivity.

However, the recognizer is far less sensitive in the second sense of the word described above. The scores provided by the recognizer are not a particularly sensitive measure of its confidence in the candidates which it has provided. The scores do not necessarily reflect the legibility of the input. For example, an incorrect word can be given a score of 100 (the maximum confidence score allowed). Likewise, it is not unusual to see a list in which the target word is ranked second, but the confidence score given to the word is lower than the score given to second ranked word in a list where the target word has been placed top. In a significant number of cases where the recognizer has ranked the intended word top, the list of alternatives is only one or two words long. It is possible for the recognizer to be more sensitive, firstly, to the legibility of the input and, secondly, to provide a better indication of its certainty (or uncertainty) in the candidates which it has generated. For example, in those cases where the recognizer has completely failed to identify the intended word then it does sometimes ascribe low confidence scores to each of the word alternatives, but this is not universal. Further work is obviously called for here.

Confidence scores support contextual selection from a list of alternatives. However, it is also possible to make use of relative confidence levels across a string of words, e.g. a sentence. Syntactic and semantic methods of disambiguation would therefore be helped by a pattern recognizer which is more sensitive to the relationship between the confidences ascribed to different lists of word alternatives.

4.4 Experiment 3: Applying the Cues

4.4.1 Introduction

Cue detection is imperfect. Alternatives values (which will be called confusions) are needed in order to derive the target word when a cue has been given the wrong value. The result of examining the list of word alternatives given by the pattern recognition software is a pattern of cues, e.g. begins with ‘c’, ends with ‘t’, contains three characters, contains an ascender and a cross, but does not contain a descender or a dot, e.g. "cat", "cot", "cut". A set of probable confusions for each of the cues is used based on the known accuracy of detection and, in the case of length, first and last, likely confusions. For instance, alternative word lengths at lower probability levels are introduced to deal with errors in the determination of this cue. A confusion matrix is used for both first and last characters reflecting the letters which are frequently confused with one another. A relatively high degree of imprecision occurs in the use of some of the cues. For example, certain letters (e.g. ‘c’, ‘e’, ‘o’) generate many possible confusions.

4.4.2 Method

The experiment used the same data set as in section 4.3. This means that catastrophic failures were excluded from the experiment. The cues which had been derived in the course of section 4.3 were used in this experiment.

Word level cues are used to derive a new list of alternatives to add to the existing list generated by the pattern recognizer. A lexicon is searched for words which match the set of cues. If the generated pattern does not occur in the lexicon then it is ignored. A list of alternatives was produced for each of the words under examination. The list of words is allowed to grow in size until a pre-determined threshold of 100 candidates is reached. The threshold is imposed to restrict consideration to the more likely confusion values.

Alternative values for each of the cues are generated. This is because cue detection is imperfect. A confusion matrix for the cues word length, first letter and last letter is constructed by using the data set and comparing the cues obtained from the data set with their target values. The confusion matrix contains information about the confusions a particular instance may have, ranked from the most likely to the least likely confusion (see Appendix N).

Word level cues are used to derive a new list of candidate words to be added to the existing list generated by the pattern recognizer. This is done in the following way. The values derived for the cues, and their alternatives, are used to search the lexicon to produce a list of words which is compatible with all the values. A ranked list of cue transformations is constructed. These transformations are constructed by comparing the cues obtained from the data set with their target values. Each transformation will generate a new pattern of cues from the initial set of cues. Each transformation is in the form 'change the value of the cue/don't change the value of the cue' for each of the seven cues.

The four binary cues (ascender, descender, dot and cross) can only have one possible confusion. A transformation which indicates that one of these four cues should change has the consequence that if the property is absent then it is set to present, whilst if the property is present then it is set to absent.

The cues first letter, last letter and word length can have many possible confusions. A transformation which indicates that one of these three cues should change also states whether the cue should be set to its first confusion, its second confusion, and so on.

The transformations are ranked by their likelihood, from the most likely transformation down to the least likely transformation. The likelihoods of the transformations were calculated by comparing the cues obtained from the data set with their target values.

For example, the most likely transformation is

> *leave all of the cues unchanged (the identity transformation)*

the next most likely transformation is

> *change the value of the length to its first (i.e. most likely) confusion, leave all of the other cues unchanged*

A less likely transformation would be

> *change the value of the length to its fourth confusion, the value of the first letter to its third confusion, change the ascender, leave all of the other cues unchanged*

The letter confusions are taken from the confusion matrices: one for first letter, and one for last letter. If a confusion does not exist then the transformation is ignored. For example, the letter 'z' only has four possible confusions so it cannot be set to its fifth confusion since one does not exist. Some of the generated patterns can be rejected because two or more of the cues are contradictory. A lexicon is searched for words which match each of the generated patterns. Patterns which do not occur in the lexicon are ignored.

Word frequency information was used to order the alternatives suggested by the word level method (see section 2.2.6). Word frequency information has been derived from a corpus of approximately 51 million words: the Oxford Corpus. The Oxford Corpus is a developmental subset of the British National Corpus (BNC). The corpus is predominantly British English. The application of word frequency has the effect of narrowing the lexical coverage. Word frequency information was applied in a very simplistic fashion; the list of word alternatives was ordered on the basis of the word frequency of each of the alternatives.

A confidence score was given to each of the alternatives generated by the word level method. The value was low (36 out of a maximum score of 100) and was decreased for each word down the list, i.e. the top ranked word was given a score of 36, the second a score of 35, and so on. The value of 36 had been chosen empirically by comparing the accuracy of the pattern recognition system with the accuracy of the word level method. Detailed results are given in Table G-1 in Appendix G. An example of the kind of output from the word level method is given in Table 4-8.

candidates generated by the pattern recognizer	confidence values generated by the pattern recognizer	candidates generated by the word level method	confidence values generated by the word level method
perhaps	34	programme	36
illustrative	34	influence	35
purposive	33	problems	34
progressive	32	production	33
provision	30	portraiture	32

Table 4-8: Sample outputs from the pattern recognizer and from using the word level method: target word 'programme'

The list of alternatives generated by the pattern recognizer was merged with the alternatives produced by the word level method. The pattern recognizer tends either to place the target word at the top of the list of alternatives, or to fail to identify the target word at all. Therefore, only the top ranked candidates from the list of alternatives are used. The top 3 candidates from the list of alternatives generated by the pattern recognizer are used [Bellaby & Evett, 1994b]. Again this threshold had been determined empirically after examining the output of the pattern recognizer. Detailed results are given in Table G-2 in Appendix G. If a word occurs both as one of the top three candidates in the list generated by the pattern recognizer and as one of the candidates generated by the word level method, then it uses the score given by the pattern recognizer. Frequency was not used to order the candidates from the pattern recognizer. The resulting merged list was ordered on the basis of the confidence score given by the pattern recognizer to each word alternative, and on the confidences of the words selected by the word level method. Table 4-9 gives an example of this.

programme
influence
problems
perhaps
illustrative
purposive
production
portraiture

Table 4-9: Merged output from the pattern recognizer and the word level method: target word 'programme'

4.4.3 Results

Table 4-10 shows the percent correct recognition after applying the word level method in those cases where the pattern recognition software had failed to identify the target. The results show a major improvement from a complete failure to provide the target word as an alternative, to producing the target word as an alternative in 50.1% of the cases (a total of 485 cases). There has been a significant change from a 100% error rate to a 49.9% error rate.

	word level method
top ranked	8.4%
top 5	26.3%
anywhere in list	50.1%

Table 4-10: Percent correct recognition using the word level method when the pattern recognizer failed to provide the target word as an alternative

Table 4-11 shows the percent correct recognition after merging the word level method with the pattern recognizer. The results are for the complete 200 word data set, excluding catastrophic failures (a total of 3,088 cases). A significant decrease in the error rate has been produced. The failure rate has dropped from 16% to 3% after the application of word level information. A small increase in the best rate has also been produced.

rank	output from recognizer	addition of word level method
1	70.9%	71.1%
2	77.7%	78.3%
3	80.0%	81.0%
4	80.9%	82.1%
5	81.6%	84.0%
6-100	84.0%	97.1%
unrecognized	16.0%	2.9%

Table 4-11: Merging the word level method with the pattern recognizer. Percent target word identified by rank

4.4.4 Discussion

It has been demonstrated that word level cues can be used to derive the target word in a significant number of cases. It is possible to use these cues to identify target words even though the pattern recognition software has completely failed to give the target word as an alternative. As with cue derivation, demonstrating the effectiveness of the approach was the main consideration at this stage. Given the demonstration of the effectiveness of the general approach, further work is necessary to develop the method and ensure that it is robust.

It should be possible to produce more accurate results. The way in which word frequency is applied is too simplistic. The integration of the pattern recognizer with the word level method can be improved. The method has not yet been proven to be robust, because the training data set is not separate from the test data set. The partial data set was used to train the method. The 200 word data set was used to test the method. However, the partial data set is a subset of the 200 word. The work presented here in Chapter 4 is intended merely to show that the word level method could be effective. However, it is clear that alternative methods should be investigated. All of these issues will be addressed in Chapter 6.

4.5 Conclusions

This experiment demonstrates that integrating word level information into the recognition process can be effective. It has proved possible to derive usable information about word level cues from the list of word alternatives given by the existing pattern recognition software. Specifically, it has been possible to use the output of the recognizer to successfully derive the target word in a significant number of cases. These results are very encouraging. This experiment is only a demonstration that the word level approach is feasible. Further work, as described in Chapter 5 and Chapter 6 should, therefore, produce better results. The experiment has shown that useful information can be gained even when the recognizer did not identify the target word. Any positive result in the case of such recognition failures would be significant, but the word level method has actually proved very successful.

The cue extraction method which has been developed has proved successful. However, it is apparent that alternative parameters and methods should be investigated. Clearly it is not possible to derive word level cues from the list of candidate words on those occasions when the pattern recognizer has experienced a catastrophic failure. Some alternative means must therefore be developed to treat such occasions. The method for deriving word level cues from the list of candidate words shows a good ability to predict word length, and is particularly accurate at the prediction of short word length. This can be useful for the identification of function words since they tend to be short. The value in identifying function words will be discussed in Chapter 7. This experiment uses information extracted from the list of alternatives generated by the recognizer. It is also possible to use pattern recognition methods to obtain this kind of information. It should be possible to increase accuracy by combining both sources of information. For instance, some of the lists of alternatives generated by the recognizer are too inaccurate for any valid information to be extracted.

The cue extraction method uses broad generalizations about character shape. This means that some inaccuracies are present. For example, one of the subjects crosses his or her z's, but the method assumes that only t's and f's are crossed, i.e. no allowance has been made for this kind of variation. Similar inaccuracies occur with ascender and descender identification. For instance, an 'f' is assumed to have both an ascender and a descender, but it is not uncommon for it to be written with just an ascender, or just a descender. It is also possible for the shape of other letters to differ from assumed letter metrics, e.g. the letter 'r' can written with an ascender and the letter 'z' can be written with a descender.

The advantages of the word level approach are that:

- it is easily applied;
- it has a low error rate;
- it deals well with poorer handwriting;
- it can often identify the target word in those cases where the pattern recognition system has failed;
- it can deal with poorly delineated characters, i.e. it bypasses some of the problems involved in segmentation;
- it is more robust than the pattern recognition system, e.g. one poorly written character can cause the recognizer to misidentify the word, this will not necessarily cause the word level approach to fail.

The disadvantages of the word level approach are that:

- it is less discriminatory than the pattern recognition system;
- it requires the support of word frequency information which has the effect of narrowing its lexical coverage.

Syntactic and semantic information can be used to make a selection from a list of alternatives. However, this is only the case if the target word occurs as one of the alternatives. The reduction in the error rate caused by the word level method is therefore of considerable benefit to the machine system. The word level method uses the information which is already present in the candidates suggested by the pattern recognizer, but structures and re-organizes this information. It also applies it in a different fashion to that of the pattern recognizer (e.g. the use of frequency information).

Word frequency information has been applied in an overly simplistic fashion. Future work will examine other ways of applying this information. A different, and probably more accurate, approach to the use of word frequency information would be to combine the candidate selection produced by the word level method with word frequency information. Word frequency should have an effect but a more reduced one than at present. Currently the effect caused by high frequency words is too great, whilst the effect of low frequency words is too small.

The pattern recognition system can be considered to be a highly discriminatory method; in those cases where it has recognized the intended word it tends to place the target word at the top of the ranked list of word alternatives. Further work is needed to make the pattern recognition system more sensitive, e.g. the relative confidence scores between different lists of alternatives and in particular the need for lower overall scores in those cases where the recognizer has completely failed to identify the intended word.

The pattern recognition system orders word candidates solely on the basis of their physical characteristics. The word level method is, by contrast, less discriminatory than the pattern recognizer but more robust. The word level method tends to generate the intended word, but requires the support of word frequency information in order to increase the probability that the intended word appears towards the top of the list. This approach bears some similarity to the human recognition framework proposed by Forster: words are chosen initially on the grounds of their physical cues but subsequently candidate sets are narrowed down and ordered on the grounds of word frequency [Forster, 1976]. The merging of the list of alternatives generated by the word level method with the list of alternatives generated by the pattern recognizer leads to a significant increase in coverage (i.e. finding the intended word) without any loss in precision (i.e. intended word top ranked), indeed there is a slight gain.

This means that a bias effect towards words that occur frequently in the written language has been introduced, but it has been restricted to confusions. In other words, physical cues prevail over frequency factors. This is not intended to be a final solution. However, the success of this method strongly suggests that this is the right approach to take. This experiment demonstrates how useful word level cues can be within the recognition process and indicates how different sources of information can be combined to good effect at the word level. Clearly, a more complex ranking procedure could lead to even better results.

Chapter 5: Improving Cue Derivation

5.1 Introduction

Chapter 4 demonstrated that the word level method is effective. The work presented in Chapter 4 showed that it is possible to use a set of word level cues to identify target words and that it is possible, indeed, to identify target words even though the pattern recognition software completely failed to give the target word as an alternative. The word level method decreases the number of target words unrecognized. It is also clear that many parameters are involved. The question is: is it possible to improve the way in which cues are derived and applied?

A method for deriving word level cues from the list of candidate words generated by the pattern recognizer, and a method for applying these cues, have been developed. These methods are presented in Chapter 4. However, during the course of this work several new variants, methods and different lines of enquiry had suggested themselves. It was therefore decided to re-examine the way in which the cues were derived. The central concern in Chapter 4 was to demonstrate the effectiveness of the approach. The present chapter attempts to improve cue derivation. The way in which data is extracted from the candidate list builds on the work presented in Chapter 4 but the approach has been modified and substantial changes made. One reason why cue derivation can be improved is that the candidate list is of no use in the case of catastrophic failures. A different source for the required information must be found. An alternative source for these cues via pattern recognition has been developed (see section 5.3). Pattern recognition is certainly a more direct, and perhaps more natural, source for these cues. Lastly, the way in which word level cues are to be applied (see Chapter 6) is also new.

The present chapter investigates two ways in which word level cues can be derived. Firstly, an improved method for deriving word level cues from the list of candidates generated by the pattern recognizer has been developed. The derivation of word level cues from a candidate list is a method which can be implemented without the need for a new pattern recognizer. The word level method can therefore be implemented easily using any existing pattern recognizer which generates a list of alternatives. Secondly, the use of a specialized pattern recognizer to obtain the data directly is examined.

5.2 Using the Candidate List

5.2.1 Introduction

There are many possible ways to estimate values for the cues. It was stated in Chapter 4 that it is possible to improve the way in which the candidate list is used to derive word level cues. The present section examines how this can be done. Several experiments are presented in this section. In these experiments the common lexicon was again used, as was the 200 word data set.

The word level method uses probabilities derived from cue detection. The prevalence of correctly identified target words in the full data set would, in consequence, mean that the observed set of initial word level cues would be given too high a probability. This is important because the main intention was to improve upon those instances in which the target did not appear or only appeared lower down the list. It is also the case that a significant proportion of the candidate lists only contain a few word alternatives, or even just one candidate. This would mean that the distorting effect of including instances in which the target word was ranked top would be further exacerbated.

The partial data set, but including catastrophic failures, was used to test direct cue extraction. Catastrophic failures could be included, in this instance, since a method to derive the cues using pattern recognition has been developed.

The full 200 word data set is used to test the word level method (see Chapter 6).

It is necessary to derive the same set of cues that were used in Chapter 4.

word length in number of character

first letter

last letter

presence or absence of ascenders

presence or absence of descenders

presence or absence of i-dots and j-dots

presence or absence of t-crosses and f-crosses

The list of alternative candidate words generated by the recognizer was examined and cue information for each of the words extracted. There are a number of different ways, variants, parameters and thresholds which have been used to extract values for the cues from the candidate list.

deriving values (see section 4.3.2)

using the arithmetic mean

using the median

using the mode

forcing a single-valued outcome (see section 4.3.3)

by rounding fractions

by reducing the list of alternatives

by choosing the value which appears highest in the list of words

fractions and bias (see section 5.2.2)

limiting the number of candidates (see section 4.3.4 and section 5.2.3)

using the ranked position of the candidates

using the confidence score of the candidates

using the difference between the confidence score of the candidates

using the ratio between the confidence score of the candidates

letter associations (see section 5.2.5)

weighting candidates by confidence score (see section 4.3.5 and section 5.2.4)

using the unadjusted confidence score

using the confidence scores raised to a set power

using normalized confidence scores raised to a set power *multiple choices (see section 5.2.6)*

Deriving values using the mean, median or mode, and forcing a single outcome by rounding fractions, by reducing the list of alternatives, or by choosing the value which appears highest in the list of words have all been described in Chapter 4. These descriptions will not be repeated here. Two ways to limit the number of candidates (using the ranked position of the candidates and using the confidence score of the candidates) and one way to weight candidates by confidence score were also described in Chapter 4. The present chapter introduces some alternative ways to implement these methods. Descriptions of these alternatives and all of the other new methods are given below.

5.2.2 Fractions and Bias

The recognizer may be more accurate at determining that a cue such as a dot is present, than it is at determining that the cue is absent. It is possible to take such a bias into account when calculating the mean average. When an average is taken then fractions need to be rounded up or down. The value at which fractions are rounded to the nearest integer can be modified. For example, the pattern recognizer showed a greater propensity to generate a cross when a cross was absent, than failing to generate a cross when a cross was present. A bias can also be caused by the words present in the lexicon because, for example, there will be a greater likelihood that the recognizer will produce candidates in which a cue is present if there are more words in the lexicon in which the cue is present than there are words in which in which the cue is absent.

One way to account for the possible existence of a bias is to modify the value at which fractions are rounded up or rounded down. For example, the cue absence/presence of descenders is a binary value. The mean average is calculated. Fractions have to be rounded up or down. Without any bias then the value at which such a fraction is rounded up or down to its nearest integer should be .5. However, a bias in the pattern recognizer or a bias resulting from the characteristics of the lexicon will be reflected in the fact that more accurate results will be obtained if this value is not .5. If a bias exists towards selecting candidates in which the cue is present then making the value at which fractions are rounded greater than .5 will compensate for the bias. Likewise, if a bias exists towards selecting candidates in which the cue is absent then making the value at which fractions are rounded smaller than .5 will compensate for this bias.

An alternative way to account for the existence of a bias is to ignore any bias effect during the process of deriving values for the cues by using the neutral intermediate value of .5 and rely on the probabilities used in the word level method to compensate for existence of any possible biases. For example, in the example list of words given above the mean for the cue absence/presence of descenders is the decimal value 0.5. If the value at which fractions are rounded to their nearest neighbour is set to .4 then this decimal value is rounded up (since .5 is greater than .4) producing a final result of 1.

The cue word length is ill suited to the first method of modifying the value at which fractions are rounded up or down. The reason for this is that an analysis of the pattern recognizer did not show a consistent bias towards either underestimating word length or overestimating word length. It is therefore better to rely on the probabilities used in the word level method to treat the range and specificalities of the biases that do exist in the performance of the pattern recognizer, e.g. the fact that the length of long words tends to be overestimated.

This first method is suited to the cues ascender, descender, dot and cross, in part because these cues are binary. However, the mean is the only method for which changing the value at which fractions are rounded up or down is applicable. For all of these cues only one of two ultimate outcomes are possible: either a 0 (indicating that the cue is absent) or a 1 (indicating that the cue is present). If mode is used and there are two outcomes with the same maximum frequency then these outcomes must be 0 and 1, producing an average of 0.5. If median is used and there are two values then these must be either 0 and 0 (producing an average of 0), 1 and 1 (producing an average of 1), or 0 and 1 (producing an average of 0.5). It is only the mean which can generate an outcome other than 0.5. Furthermore, if the cutoff is 0.5 then mean, median and mode will all produce the same result. In the case of the cues ascender, descender, dot and cross the use of a bias can be seen as another way to force a single integer outcome because a result of .5 will occur when the number of candidate words in which the property is absent is equal to the number of words in which the property is present.

More accurate results can be produced by taking bias effects into account during the process of calculation. However, it should also be noted that this method might distort the information extracted since some legitimate indicators are ignored. Furthermore, it can destroy potentially useful information such as the inter-relationship between certain characters and the cues ascender, descender, dot and cross, e.g. the fact that a word beginning with the letter 'd' should have at least one ascender present.

5.2.3 Number of Candidates

The work presented in Chapter 4 suggested that limiting the size of the candidate list was an effective approach. However, the two methods used displayed different, but perhaps complementary, characteristics. It may be is possible to use alternative methods which combine aspects of the two methods given in Chapter 4.

Four different ways to limit the number of candidate words were considered.

1) A threshold based on the ranked position of the candidates.

2) A threshold based on the confidence score of the candidates.

3) A threshold based on the difference between the confidence score of the top ranked candidate and the confidence scores of the other candidates in the list.

4) A threshold based on the ratio between the confidence score of the top ranked candidate and the confidence scores of the other candidates in the list.

5.2.4 Weighting by Confidence Score

Limiting the size of the candidate list may be too blunt an instrument. For example, words further down the list bear less resemblance to the target word, but they do resemble the target word.

The confidence scores used by the pattern recognizer have no formal statistical significance. An adjusted set of scores may increase the accuracy of cue derivation. Raising the confidence score of candidates to a power greater than 1 increases the influence of candidates with higher scores. For instance, if the confidence scores are raised to the power 2.8 then, whilst none of the candidates are explicitly ruled out of bounds, the effect of smaller confidence scores is decreased and therefore the influence of candidates lower down the list of alternatives is less than the influence of candidates higher in the list.

Raising the confidence score of candidates to a power smaller than 1 decreases the influence of candidates with higher scores. In this case, the effect of higher confidence scores is decreased and therefore the influence of candidates higher up the list of alternatives decreases.

Whether the confidence score of candidates is raised to power higher or smaller than 1, the advantage which weighting has over limiting the size of the candidate list is that all of the candidates retain an influence on the final result.

The top ranked candidate does not always have a confidence score of 100, or even close to 100. The lowest confidence score given by the pattern recognizer to a top ranked candidate in the 200 word data set is 27. The influence of higher ranking candidates will be greater in word lists where higher ranking candidates have a large confidence score, than in word lists where higher ranking candidates have a small confidence score. It may be more useful to increase the influence of higher ranking candidates in all of the lists to a similar degree. This can be done by normalizing the confidence scores.

Three ways to use the confidence score have been considered.

1) Using the unadjusted confidence score.

2) Using the confidence score of each candidate raised to a set power.

3) Using the confidence scores of the candidates so that they are first normalized and subsequently raised to a set power. A scaling factor S is calculated by

dividing 100 by the confidence score of the top ranked candidate. The confidence score of each candidate is then multiplied by S. The top ranked candidate will end up with a score of 100 (the maximum confidence score allowable). After normalization, the confidence score of each candidate is raised to a set power [Ho, et. al., 1992b; Huang & Suen, 1993].

For example, consider the example list of word alternatives given in Table 5-1. Table 5-1 shows the list of candidates, their word length, the confidence score given to each candidate by the pattern recognizer, the confidence scores raised to the power 2.9, and the confidence scores normalized and subsequently raised to the power 2.9. The value 2.9 has been chosen just for the sake of example.

candidate	word length	confidence score	confidence score raised to an example value of 2.9	normalized confidence score raised to an example value of 2.9
four	4	90	464,840	630,957
tour	4	59	136,606	189,094
low	3	55	111,443	150,472
tom	3	48	75,093	100,092
for	3	47	70,645	94,713

Table 5-1: Example list of candidates including confidence scores

The mode for the cue word length would be calculated using the method in section 4.3.2 as follows. It can be seen that 3 occurs three times and 4 occurs twice. The mode is therefore 3.

If the confidence score is taken into account then 3 occurs 150 times (55 + 48 + 47) and 4 occurs 149 times (90 + 59) and the mode is therefore 3.

If the confidence score raised to the power 2.9 is used then 3 occurs 257,181 times (111,443 + 75,093 + 70,645) and 4 occurs 601,446 times (464,840 + 136,606) and the mode is therefore 4.

If the adjusted confidence score raised to the power 2.9 is used then 3 occurs 345,277 times (150,472 + 100,092 + 94,713) and 4 occurs 820,051 times (630,957 + 189,094) and the mode is therefore 4.

All four of the different ways to limit the number of candidate words given above can be combined with weighting by confidence score. The use of weighting can be considered to be an implicit way to limit the number of candidate words. For instance, if the confidence scores are raised to the power 2.9 then, whilst none of the candidates are explicitly ruled out of bounds, the effect of smaller confidence scores is decreased and therefore the influence of candidates lower down the list of alternatives is swamped by the influence of candidates higher in the list.

Two contrary impulses are apparent in the use of weighting: increasing the importance of candidates which are higher ranked (and hence bear a greater degree of resemblance to the target word) whilst not swamping the influence of lower ranked words and the potentially useful information which they contain.

5.2.5 Letter Associations

Letter confusions tend to share common properties of shape with the intended letter, i.e. human readers tend to confuse the group of short letters which are rounded on the left (‘e’, ‘o’, ‘c’) with each another more often than with dissimilar letters [Bouma, 1971]. Neural network models exploit the notion that a single letter cue (e.g. rounded on the left) activates all of the letters which contain this cue [e.g. McClelland & Rumelhart, 1981; Rumelhart & McClelland, 1982]. In a very simplistic fashion this notion is also taken up here.

In order to exploit the assumed confusions produced by letters which share common properties of shape an instance of a letter in the candidate list increases the score for the letter and also increases the score of associated letters. This approach can help to resolve ambiguous results. For example, it is possible that mode may not give a clear measure of central tendency. For instance, if the letter 'c' occurs twice, the letter 't' occurs twice, and the letter 'a' occurs once then it is not clear whether 'c' or 't' is the mode. However, the letter 'a' appears to share common properties of shape with the letter 'c' whilst resemblance is less apparent between 'a' and the letter 't', and this would suggest that the letter 'c' is more clearly indicated than the letter 't'.

The mode is weighted using the confidence score of the candidates as described previously, so that an instance of a letter increases the observed frequency of that letter by the confidence score (or by a function of the confidence score) of the candidate in which it was observed. However, an instance of a letter also increases the observed frequency of the letters with which it is typically confused.

A confusion matrix for both of the cues first letter and last letter was constructed by using the complete data set and comparing the cues obtained from this data set with their values. The confusion matrix contains information about confusions a particular instance may have, together with the probability of this confusion occurring. Appendix H contains these two confusion matrices. The letter confusions and their associated probabilities were derived from an analysis of the complete data. The complete data set can be considered to be large: a total of 3,082 target words and 42,500 candidate words, and hence an adequate representation of the relationships between individual letter cues made by the pattern recognizer.

The amount associated letters are increased is proportional to the amount that the original letter was increased, to the degree of resemblance between the letters, and to a constant. This constant was different for letters which had already been suggested as possible candidates and those which had not (see Appendix H). The confusion probabilities in the confusion matrix were used to calculate the increase in score for each of the associated letters (see Appendix H).

5.2.6 Multiple Choices

It is possible to use the mode to derive more than one alternative for the cues first, last and length. It is not useful to do this with the cues ascender, descender, dot and cross because they are binary values. The use of multiple choices is particularly relevant when two or more alternatives have the same maximum frequency, which would suggest that, at least without additional processing (e.g. reducing the candidate list), they are all equally likely. The use of multiple choices is also used to indicate that a further likely choice is possible.

The algorithm used to derive a first and a second choice is as follows:

1) If there is only one candidate, make this the first choice and leave the second choice empty.

2) If there is the only one category that occurs with greatest frequency and only one category that occurs with the second greatest frequency, then make the former the first choice and the latter the second choice

3) If there are two categories which both share the same greatest frequency then record them both and reduce the candidate list by progressively removing its lowest ranked candidate until only one of the two recorded categories occurs with the greatest frequency, i.e. the other is reduced. Make the category that now occurs with the greatest frequency the first choice and make the other recorded category the second choice.

4) Otherwise, reduce the list of alternatives by removing its lowest ranked candidate and return to step 1.

This algorithm can be easily extended for three or more choices.

An advantage of implementing multiple choices is that the confusion matrix may not include the correct confusion. For example, the data set used in these experiments does not include a word beginning with the letter 'x'. There is always going to be the possibility that a confusion matrix may not be complete, or that certain confusions are so unlikely that they have a very low probability rating. The use of multiple choices can allow a confusion to be generated that is not in the confusion matrix, or to boost the chances of selection of an unlikely confusion.

5.2.7 Experiments to Establish Parameters

5.2.7.1 Introduction

Various different approaches to the derivation of word level cues using the candidate list and variants upon these approaches have been set out.

1) There are three basic methods for deriving the information: mean, median and mode (see section 4.3.2).

2) The median and the mode can produce more than one outcome. However, only one value is desired. It is possible to:

a) take the average of multiple outcomes.

b) reduce the list until only one value is generated.

c) pick the value which occurs first in the list of alternatives.

Taking the average is only appropriate to the cues length, ascender, descender, dot and cross. The cues first, last and length are only suited to reduction and picking (see section 4.3.3).

3) When an average is taken then fractions need to be rounded up or down. The value at which fractions are rounded to the nearest integer can be modified. The obvious cutoff point for rounding is the mid-way point (.5), so that fractions smaller than .5 are rounded down, and fractions greater than .5 are rounded up. The mathematical convention for rounding is used when the fraction is exactly .5 (see section 5.2.2).

4) Since lower ranked words should bear less resemblance to the target word then an explicit threshold below which candidate words are ignored, or an implicit threshold which promotes the importance of candidates higher up the list of alternatives, will improve results (see section 4.3.4 and section 5.2.3).

5) It is possible to weight the methods by taking the confidence scores ascribed to each candidate word into account (see section 4.3.5 and section 5.2.4).

6) Letter associations can be used so that an instance of a letter in the candidate list increases the score for the letter and also increases the score of associated letters (see section 5.2.5).

7) A first and a second choice can be derived for the cues first, last and length when the mode is used (see section 5.2.6).

The aim of the present investigation is to compare ways of deriving values. A number of different methods and variants have been described. It was pointed out in Chapter 4 that the large number of combinations of the different alternatives means that it is not practical to provide results for each and every one of the alternatives. The number of methods used in the present approach is greater than in Chapter 4. Furthermore, the cutoff point for rounding fractions, the threshold used to limit the number of candidates, and the amount associated letters are increased can all take a range of values. This only serves to increase the number of possible combinations. Cue derivation involves estimating values, estimating probabilities, setting thresholds, combining values, etc. Clearly, many parameters are involved.

The evaluation criteria used were different from the criteria described in Chapter 4. Two criteria were used:

1) accuracy of cue detection

2) best combination of methods

It was apparent from examining the results of combinations of different methods of obtaining word level cues that the best results were obtained when the combination minimized dissimilarity, where dissimilarity is taken to be a weak resemblance between the pattern and the target (see Appendix I). The best combination of methods minimized the likelihood of any of the values being incorrect when taken together. Reasons for this are that the optimum point for one method might not be the same as the optimum point for a second. In particular, a method might be able to extract useful information from one list of word alternatives whilst a second method might not. A simple replication of the output of the pattern recognizer was to be avoided. Cue derivation has to generate new information if the word level method is to generate new word alternatives.

It was not feasible to test all of the methods for deriving word level information from the candidate list together because of the great variety of possible methods, and indeed variations of these methods. The large number of possible alternatives made it impracticable to construct one single test.

Some of the different methods can be investigated on their own. It is also useful to compare some of the methods directly. These are:

1. the cutoff point for rounding fractions;

2. a comparison between unweighted and weighted;

3. the use of letter associations;

4. the extraction of both a first and second choice.

These investigations helped to establish the parameters within which a final choice could be made.

Whenever results are given for a particular method, or methods, the combination of alternatives that produced the most accurate results for all the other methods have been used. In order to make the presentation of results neater, only selected results have been given in this chapter. Detailed results have, however, been given in the Appendices.

5.2.7.2 Experiment 4a: Fractions and Bias

Modifying the cutoff point for rounding up or down can serve to compensate for the existence of a bias in a recognizer which causes it to unduly favour the absence, or the presence, of a cue.

A test was carried out to see what effect different cutoff points for rounding up or down made to the accuracy of cue detection. Cue distribution is not uniform in the data set. This could cause results to be biased by the data set Four sets of test data were therefore created, one for each of the cues to be examined. For each test data set, targets were selected from the complete data set using a random procedure. In each case, the number of targets in which a cue was present was made equal to the number of targets in which a cue was absent. Each test data set was therefore balanced with respect to the test criteria, i.e. the absence or presence of the cue currently under examination. This resulted in a data set of 250 words for ascender, 562 words for descender, 532 words for dot, and 604 words for cross. Table 5-2 shows the effect of a range of bias values on cue derivation for the cues ascender, descender, dot and cross. Results are given for percent correct with bias correction.

cutoff	ascender	descender	dot	cross
0.1	52.4%	63.5%	55.6%	54.8%
0.2	54.0%	69.2%	60.5%	59.4%
0.3	55.2%	72.6%	67.7%	63.6%
0.4	58.0%	74.2%	73.7%	67.7%
0.5	62.0%	71.4%	78.0%	68.4%
0.6	61.2%	70.3%	81.2%	68.2%
0.7	61.2%	67.6%	82.0%	67.4%
0.8	61.2%	64.2%	82.0%	65.2%
0.9	61.2%	62.3%	81.2%	62.6%

Table 5-2: Percent correct of the cues ascender, descender, dot and cross for a range of bias values

Table 5-3 shows the bias correction which produced the most accurate results. Results are given for percent correct both without and with bias correction, and the bias value which produced the most accurate results.

cue	percent correct without bias correction	most accurate bias value	percent correct with bias correction
ascender	62.0%	.5	62.0%
descender	71.4%	.4	74.2%
dot	78.0%	.7	82.0%
cross	68.4%	.5	68.4%

Table 5-3: The cues ascender, descender, dot and cross both without and with bias correction

A clear pattern can be seen for the cues descender and dot. The results suggest that a bias exists in the recognizer in its recognition of both of these cues. The recognizer is failing to report descenders when they exist in the target, and reporting dots when none exist in the target. However, no such clear pattern could be seen for the cues ascender and cross. This suggests that the recognizer is not biased in its determination of ascender presence/absence or of cross presence/absence.

5.2.7.3 Experiment 4c: Letter Associations

The use of letter associations increased the number of first and last letters correctly identified. Table 5-4 compares the percent correct of the cues first letter and last letter without and with the use of letter associations.

method	first letter, percent correct without letter associations	first letter, percent correct with letter associations	last letter, percent correct without letter associations	last letter, percent correct with letter associations
rank, mode, reduce	36.1%	42.7%	42.0%	47.6%
rank, mode, initial	36.1%	42.7%	42.0%	47.6%
score, mode, reduce	36.6%	41.8%	43.0%	46.5%
score, mode, initial	36.6%	42.0%	43.0%	46.5%
difference, mode, reduce	37.1%	42.5%	41.7%	47.4%
difference, mode, initial	37.1%	42.5%	41.7%	47.4%
ratio, mode, reduce	37.2%	42.8%	42.2%	47.4%
ratio, mode, initial	37.4%	42.8%	42.3%	47.4%
power, mode, reduce	37.6%	42.3%	41.8%	48.3%
power, mode, initial	37.6%	42.3%	41.8%	48.3%
power + normalization, mode, reduce	37.6%	42.5%	41.8%	48.6%
power + normalization, mode, initial	37.6%	42.5%	41.8%	48.6%

Table 5-4: Percent correct both without and with letter associations: partial data set

The use of letter associations increased the number of first and last letters correctly identified for all the different combinations of alternatives, e.g. accuracy increased whichever approach towards limiting the number of candidates was used, etc. Naturally, the degree of improvement gained by the use of letter associations differed between the different combinations.

5.2.7.4 Experiment 4b: Weighting

Weighting the methods by the confidence scores was, in general, the best approach for all of the methods. A threshold was firstly selected for each of the methods used to limit the number of candidates or adjust the confidence scores of the candidates. The threshold which, on average, was the most accurate for all of the cues combined was selected. Detailed results are given in Appendix J. Table 5-5, Table 5-6 and Table 5-7 give the results for just the best threshold.

Table 5-5 presents results for limiting the number of candidates by their score (score), by their rank (rank), by the difference between the confidence score of the top ranked candidate and the confidence scores of the other candidates (difference), and by the ratio between the confidence score of the top ranked candidate and the confidence scores of the other candidates (ratio) when weighting is not used. For all of the other methods, the combination of alternatives that produced the most accurate results for a given cue have been used. Results are for the partial data set. The last column of the table shows the value of the threshold used to limit the number of candidates.

method	length	ascender	descender	dot	cross	first	last	threshold
rank	37.2%	80.9%	72.0%	79.7%	68.0%	38.4%	42.0%	8
score	37.2%	80.2%	72.2%	80.1%	69.0%	35.9%	43.2%	30
difference	38.9%	81.7%	71.5%	79.2%	68.4%	37.2%	42.2%	42
ratio	39.2%	80.7%	72.7%	80.1%	67.5%	37.6%	41.4%	73

Table 5-5: Percent correct of the cues using the four methods of limiting the number of candidates when unweighted: partial data set

Results for rank, score, difference and ratio when weighting is used are also given in Appendix J. Results for the best threshold given in Table 5-6. The last column of the table shows the value of the threshold used to limit the number of candidates.

method	length	ascender	descender	dot	cross	first	last	threshold
rank	38.4%	81.5%	72.3%	80.2%	69.0%	42.7%	47.6%	16
score	38.9%	80.9%	72.0%	79.6%	69.4%	41.8%	46.5%	35
difference	38.9%	82.5%	72.3%	79.4%	68.5%	42.5%	47.4%	44
ratio	38.9%	81.9%	71.0%	79.2%	68.4%	42.8%	47.4%	53

Table 5-6: Percent correct of the cues using the four methods of limiting the number of candidates when weighted: partial data set

Detailed results for weighting by the confidence scores raised to a set power (power) and for weighting by the normalized confidence scores raised to a set power (power plus normalization) are provided in Appendix J. Results obtained for the best threshold are given in Table 5-7. The last column of the table shows the power to which the confidence scores are raised.

method	length	ascender	descender	dot	cross	first	last	threshold
power	39.5%	82.4%	72.2%	80.4%	68.7%	42.3%	48.3%	2.2
power + normalization	39.0%	82.7%	73.3%	79.7%	68.4%	42.5%	48.6%	2.9

Table 5-7: Percent correct of the cues using the two methods of adjusting the confidence scores: partial data set

After the best threshold was selected for each of the methods used to limit the number of candidates or adjust the confidence scores, the best combination of other methods for each cue was chosen. Appendix J again gives the detailed results. Results are only shown for the selected threshold. However, results for all of the other combination of methods are provided. Results are for the partial data set.

There was no strong indication that one method consistently favoured any particular cue. It was decided to use the method which, on average, was the most accurate for all of the cues combined. Likewise, it was decided to use the particular level of threshold which was, on average, the most accurate for a method for all of the seven cues.

When the confidence score of the candidates was not used, rank and score appeared less accurate than the other methods. The difference method was more accurate than rank and score, but not as accurate as ratio. The ratio method appeared to be slightly more accurate than the other methods.

When weighting was used, rank and score could be rejected as obviously inferior to the other methods. The difference method was more accurate than rank and score. The ratio method, power and power plus normalization all produced a similar degree of accuracy, and the distribution of these results for each of the seven cues was substantially the same. However, power plus normalization performed slightly better overall than the other two methods.

Weighting by the confidence score typically produced results which were more accurate than those which were unweighted. This indicates that the confidence scores ascribed by the pattern recognizer do, at least to some extent, reflect the resemblance of the candidate to the target word. However, the increase in accuracy was not very strong and this suggests that the confidences scores generated by the pattern recognizer are not as sensitive to the relative differences between candidates as would be desirable. Furthermore, the increase in accuracy caused by weighting the mean, median and mode was not equal for all of the cues and, in the case of the cue dot, a decrease in accuracy occurred. However, the largest comparative loss by using weighting (on the cue dot) was smaller than the largest comparative gain on all of the other cues.

5.2.7.5 Experiment 4d: Multiple Choices

A precise evaluation of the use of multiple choices is difficult. This is because a proportion of the cues will not have a second choice, or if they have a second choice, will not have a third choice, and so on. If a further choice is not available, the confusion matrix is used instead. However, in practice, a proportion of the alternatives in the confusion matrix will be the same as the choices already used.

Whilst it is not possible to make a precise statement about the efficacy of multiple choices, it is possible to compare the accuracy of multiple choices with the accuracy of a single choice combined with values obtained from the confusion matrices (see section 6.2.2). Such a comparison will provide a good indication of the relative worth of the methods.

Results for extracting multiple choices using the mode for the cues word length, first letter and last letter are given in Table 5-8. The table shows the percent correct of the cues for: first choice; first and second choice combined; first, second and third choice combined; first, second, third and fourth choice combined; and first, second, third, fourth and fifth choice combined. Table 5-8 also shows results for using a single choice combined with values obtained from the confusion matrices. The table shows the percent correct of the cues for: single choice; single choice combined with the first alternative from the confusion matrix; single choice combined with the first and second alternative; single choice combined with the first, second and third alternative; single choice combined with the first, second, third and fourth alternative.

method	word length	first letter	last letter	average
first choice	38.6%	42.2%	47.4%	42.7%
second choice	58.5%	55.2%	63.5%	59.0%
third choice	68.5%	58.2%	64.7%	63.8%
fourth choice	72.8%	60.3%	66.4%	66.5%
fifth choice	73.6%	61.1%	66.7%	67.2%
single choice	39.0%	42.5%	48.6%	43.4%
first confusion	67.2%	53.9%	64.9%	62.0%
second confusion	82.5%	61.9%	74.0%	72.8%
third confusion	88.6%	68.4%	79.1%	78.7%
fourth confusion	93.4%	73.6%	84.5%	83.9%

Table 5-8: Percent correct for multiple choices compared to a single choice combined with the confusion matrices: partial data set

The only instance where multiple choice is better than a single choice is the cue first letter, for first and second choice combined. In all the other instances (including the average for first and second choice combined), the accuracy of cue detection for multiple choices is worse than for a single choice combined with the confusion matrix. The results suggest that the use of a single choice will be a better approach than use of multiple choices.

5.2.7.6 Overall Results

Two types of control were used in these tests. Firstly, the cues of the top ranked candidate were used a comparative control in order to demonstrate that additional information was indeed being supplied by examining the candidate list. Secondly, values derived from an examination of the whole candidate list were used as a control to show that the imposition of a threshold did lead to greater accuracy.

The best general approach is weighting by the normalized confidence scores raised to the power 2.9. It has proven more effective to weight the cues according to the confidence score of their source words than to leave them unweighted. Limiting the number of candidates has been rejected in increasing the influence of candidates with higher confidence scores. The method which has been chosen does not produce the absolute best result for each and every one of the cues. For example, limiting the number of candidates by the ratio between the confidence score of the top ranked candidate and the confidence scores of the other candidates produced the best results for the cue word length. The selected method should demonstrate consistently good results for all of the cues. Weighting by the normalized confidence scores raised to a set power produced the best results overall.

Results for the best approach are given in Table 5-9. Results for using a first and second choice have been set aside for the reasons given above.

method	length	ascender	descender	dot	cross	first	last	threshold
power + normalization	39.0%	82.7%	73.3%	79.7%	68.4%	42.5%	48.6%	2.9

Table 5-9: Percent correct for the best combination of methods

The evaluation criteria described previously were used to select the best combination of methods, variants and thresholds. The values were derived in the following ways:

word length: the median, taking the average of multiple outcomes to force a single integer outcome.

first letter: the mode, reducing the list of alternatives until one most frequent value is left.

last letter: the mode, reducing the list of alternatives until one most frequent value remains.

ascenders: the mean, rounded down at 0.5.

descenders: the mean, rounded up at 0.4.

dots: the mean, rounded up at 0.7.

crosses: the mean, rounded down at 0.5.

A partial data set was used for testing in the present investigation. In order to compare the new approach with the approach described in Chapter 4, results need to be given for the data set used in Chapter 4. Table 5-10 compares the method for deriving word level cues from the list of candidate words which was presented in Chapter 4 with the new method. Results are for the complete 200 word data set, including uninformative word lists, but excluding catastrophic failures (a total of 3,088 cases).

method	length	ascender	descender	dot	cross	first	last
approach described in Chapter 4	70.5%	91.1%	88.9%	89.4%	89.1%	71.9%	75.4%
approach described in Chapter 5	76.2%	93.4%	92.5%	94.5%	91.1%	78.7%	80.2%

Table 5-10: Comparison between the approach described in Chapter 4 and the approach described in Chapter 5

Results for the new method gave an increase in accuracy for all of the cues. It has therefore proved possible to improve the way in which cues are derived from the candidate list.

5.2.7.7 Discussion

An improvement in the way in which cues are derived from the candidate list has been demonstrated. Cue derivation has been based on the output of one pattern recognizer. Values would have to be calculated for any particular recognizer, but the method could be used by any recognizer which generates a list of alternatives.

The investigation of cue derivation involves many parameters and it is a complicated task to evaluate them in conjunction with one another. The values used in the present work have been derived empirically and reflect the biases of the pattern recognizer being used. Given the demonstration of the effectiveness of the cues, further work is necessary to develop some way of optimizing the parameters both to improve the effectiveness of the method and to facilitate its application to other recognizers.

The pattern recognizer attempts to place the target word at the top of a ranked list of alternatives. It will be possible to use the methods described above with most pattern recognizers. However, the degree to which the particular scoring method adopted reflects a candidate's resemblance to the target word will vary. Likewise, the characteristics of other pattern recognizers will not be identical to the characteristics of the particular pattern recognizer which has been used in this research. The general approach described in this chapter will, therefore, be of use to any handwriting recognition system, although the particular strategy that will be the most appropriate for an individual system will vary.

Words further down the list bear less resemblance to the target word, so various thresholds for the maximum number of words were tested. The score given by the recognizer to each word is meant to indicate how confident the recognizer is that the candidate resembles the target word. The usefulness of applying this information to the values produced by cue extraction was also tested. Weighting the cues according to the confidence score has proven more effective than leaving them unweighted.

An explicit threshold which caused candidates lower down the list of alternatives to be ignored proved to be least successful. It was determined that the most accurate results, on average, were obtained if candidates were weighted by their confidence score raised to a set power, after the confidence scores of the list of alternatives had first been normalized.

The use of a ratio between the confidence scores also proved successful. However, a ratio between the confidence score of the top ranked candidate and the confidence scores of the other candidates has similar mathematical properties to raising the confidence score of each of the candidates to a set power.

The confidence scores generated by the pattern recognizer are not as sensitive as would be desirable. This can be seen by looking at the differences between the unweighted and weighted methods. Weighting the instances by the confidence score of their source word is more accurate than not weighting them. However, the improvement was not as strong as one would expect if the pattern recognizer was indeed responding strongly to the degree of resemblance between the proposed candidate and the target word. It might be the case that the strength of the response is being swamped by the recognizer's need to consider the other letters in the word.

The extent to which the pattern recognizer is able to identify those candidates which most resemble the target is demonstrated by the degree, if at all, that weighting affects the accuracy of cue detection. These results provide a useful insight into some the characteristics of the pattern recognizer. These results show that the recognizer is more accurate at determining some cues (e.g. length, ascender and first) than it is at determining other cues (e.g. descender, cross, and last), and very poor at determining the cue dot. The sole reason for the difference in accuracy between the unweighted and weighted results is the confidence score which the recognizer ascribes to its proposed candidates. The results therefore show that the recognizer is not using its information about dots to its full potential. Setting aside the use of weighting for the purpose of deriving values from the candidate list for the word level method, the kind of comparison which has been set out here can also be seen to be a useful way in which to test the ability of a pattern recognizer to recognize a particular cue. It would be useful to cross-reference such results with some other measure of sensitivity or accuracy.

For example, an examination of direct cue extraction (see section 4.3) also suggests that the pattern recognizer is failing to recognize dots. This suggests that the increase given to the confidence values of the characters 'i' and 'j' when a dot is found is not high enough. The recognizer is also (perhaps as a corollary of this) misrecognizing other characters as being an `i' or a `j' too frequently.

5.3 Experiment 5: Direct Cue Extraction

5.3.1 Introduction

An alternative source for word level cues is direct cue extraction. That is the use of pattern recognition methods to recognize the cues from the input data. Direct cue extraction is perhaps a more natural source for these cues. Direct cue extraction certainly improves upon the way in which word level cues were derived in Chapter 4 and in section 5.2 because it means that the word level method can be applied even on those occasions when the pattern recognizer fails to generate any candidates. The information extracted from the list of alternatives generated by the recognizer can also be merged with the information obtained from direct cue extraction.

5.3.2 Method

The pattern recognition methods used in this experiment were developed for a word shape recognizer [Powalka, et. al., 1994; Powalka, 1995]. An example of output from the word shape recognizer is given in Table 5-11. Although the same underlying information is used by the word shape recognizer and by direct cue extraction, the way in which this information is interpreted and subsequently employed is not the same. The differences between the two approaches are examined in section 6.7. It was not possible to use the word shape recognizer to derive values for the cues first and last since the word shape recognizer uses a post-detection letter verification technique.

number of vertical bars	14
middle zone probability	22.0
middle and upper zone probability	96.1
middle and lower zone probability	0.0
upper, middle and lower zone probability	33.0
number of dots	1
number of dashes	1

Table 5-11: An example of output from the word shape recognizer

The word shape recognizer attempts to identify ascender and descender sequences; their existence and their position within the physical word. It describes the input using three zones: upper, middle and lower. It calculates the probability that the input data lies within these zones and its position within the sequence of ascenders and descenders. In practice, four sets of information are used: the probability that data lies solely within the middle zone, the probability that it lies within both the middle and the upper zone, the probability that it lies within both the middle and the lower zone, and finally the probability that it lies each of these zones. The word length is calculated from the number of times an imaginary line through the middle of a word crosses the path of the pen strokes. The position of ascenders and descenders is expressed in terms of the number of such intersections. Dot and dashes are only used in conjunction with middle zone information.

The probability scores for the zones were examined and the highest score used to decide whether the presence of an ascender was indicated, and whether the presence of a descender was indicated. The actual number and positioning of the cues was ignored since such information is irrelevant for the purposes of this experiment. The number of dots and dashes were examined in order to produce the same kind of information about the likely presence of dots and crosses.

This experiment used the vertical bar recognizer in order to calculate word length (expressed as the number of characters in the word). The vertical bar recognizer attempts to identify the number of approximately vertical strokes which are directed downwards.

The average number of letters per vertical bar can be calculated. For example, the letter 'm' is typically written using 3 vertical bars, 'l' with 1 vertical bar, and 'k' using 2 vertical bars. The average number of bars per letter is 1.4. The average number of letters per bar is 0.7. The estimated number of vertical bars in the input data divided by the number of bars per letter gives the length of the word. Fractions are, for simplicity, rounded to their nearest integer value.

Catastrophic failures occur when the pattern recognizer has generated no candidates at all. This means that it is not possible to use the candidate list to derive word level cues. Direct cue extraction can be used to extract only five of the seven word level cues (i.e. word length, ascender, descender, dot and cross). The word shape recognizer cannot derive values for the cues first and last, but it is possible to use the pattern recognizer to obtain this information. Whilst it has proved impossible for the pattern recognizer to match the letter strings it has generated with any of the words in its lexicon, it is possible to examine the characters recognized by the pattern recognizer in order to extract the remaining two cues (first and last). The pattern recognizer was used to extract character alternatives from the ink data. Each and every character recognized by the pattern recognizer, together with the physical location of the character, was recorded. The x-y co-ordinates of the character's start point and end point were used. The confidence score given by the pattern recognizer to each character alternative was also obtained. The most likely candidate for the first and last letter is calculated from this data.

5.3.3 Results

Results for three different data samples are given here. Firstly, results for the partial data set (the data set used in Experiment 4) are provided. A second data set was also used: those words where the pattern recognizer suffered a catastrophic failure. It is possible to use direct cue extraction to derive word level cues when a catastrophic failure has occurred. Lastly, results are provided for the complete data set, except for those target words where the pattern recognizer suffered a catastrophic failure. Note that this data set therefore includes the partial data set, but excludes the catastrophic data set.

Table 5-12 shows percent correct recognition of the cues length, ascender, descender, dot and cross using direct cue extraction using the partial data set.

cue	percent correct using the candidate list	percent correct using direct cue extraction
word length	39.0%	40.5%
ascender presence/absence	82.7%	87.8%
descender presence/absence	73.3%	93.4%
dot presence/absence	79.7%	90.6%
cross presence/absence	68.4%	67.2%

Table 5-12: Percent correct recognition using the candidate list and using direct cue extraction: partial data set

The results of deriving word level information from the candidate list and of using direct cue extraction are comparable. Direct cue extraction had slightly better results than using the candidate list with respect to ascenders, and was better than using the candidate list with respect to both descenders and dots. Direct cue extraction was worse than using the candidate list with respect to crosses. Direct cue extraction correctly identified the word length more often than using the candidate list. However, length information derived from the candidate lists had a lower margin of error than direct cue extraction. See Figure 5-1.

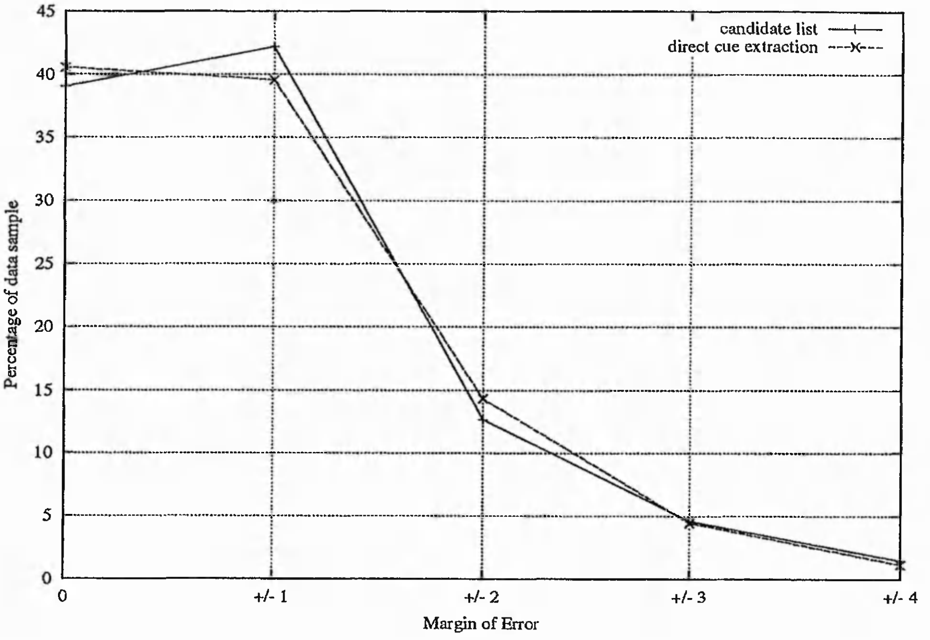


Figure 5-1: Comparison between direct cue extraction and using the candidate list for word length: partial data set

Table 5-13 shows percent correct recognition for all of the cues in those cases where the pattern recognizer suffered a catastrophic failure.

cue	percent correct
word length	24.2%
ascender presence/absence	93.9%
descender presence/absence	93.7%
dot presence/absence	87.5%
cross presence/absence	62.7%
first letter	26.6%
last letter	28.0%

Table 5-13: Percent correct recognition using direct cue extraction for catastrophic failures

Table 5-14 compares the performance of the two sources of word level cues for all of the cases where the pattern recognizer managed to generate a list of word alternatives. Only those cases where the pattern recognizer suffered a catastrophic failure have been excluded from this table.

cue	percent correct using the candidate list	percent correct using direct cue extraction
word length	91.1%	70.7%
ascender presence/absence	76.2%	39.2%
descender presence/absence	93.4%	87.8%
dot presence/absence	92.5%	93.2%
cross presence/absence	94.5%	93.9%

Table 5-14: Comparison between using the candidate list and direct cue extraction: complete data set, excluding catastrophic failures

When direct cue extraction and using the candidate list are compared for the complete data set (excluding catastrophic failures) then using the candidate list is better than direct cue extraction for all of the cues except dots. However, the difference between the two methods is not great except in the case of the cue word length and, to a lesser extent, the cue ascenders. The reason for the differences between Table 5-12 and Table 5-14 is that the former is for the partial data set, whilst the latter is for the complete data set, excluding catastrophic failures. Information extracted from the candidate list will be more accurate if the target is present, particularly if the target is ranked high.

5.3.4 Discussion

It has proved possible to use the data produced by the word shape recognizer in order to generate some of the word level cues. These results indicate that, for the partial data set, direct cue extraction is more accurate than using the candidate list on many of the cues, a notable exception being the cue cross. However, the differences between the two methods are not great. Some of the lists of alternatives generated by the recognizer are too inaccurate for any valid information to be extracted. The pattern recognizer's need to recognize all the characters in a word, and the inevitable whole word recognition errors which this causes, may be the reason why direct cue extraction is sometimes more accurate than using the candidate list.

Whilst a specialized word-shape recognizer has been used to derive some of these word level cues, it is apparent that it is possible to modify an existing conventional character-based recognizer to perform the same task, e.g. to calculate the number of vertical strokes present within the input data. Whilst some of the same underlying information is used by both methods, the word shape recognizer uses this information in a different way to that of the word level method.

The advantage of using the candidate list in order to derive word level cues is that it is easily applied. Any pattern recognizer which produces a list of word alternatives could be used in a similar manner to derive these cues. The level of accuracy of direct cue extraction and of using the candidate list is comparable. This means that it is possible to derive word level cues easily and quickly using any conventional pattern recognizer and so implement the word level method without the need for additional pattern recognition methods. It is only in those cases where a pattern recognizer has failed to generate any word alternatives that it becomes necessary to resort to additional pattern recognition methods of analysis.

Word level cues can be used as a replacement for the pattern recognizer in those cases where the pattern recognizer has experienced a catastrophic failure. This is because it has proved possible to use the pattern recognizer to generate the cues first and last, albeit that cue detection is not particularly accurate for these two cues (26.59% and 27.98% respectively).

5.4 Experiment 6: Merging the Two Approaches

5.4.1 Introduction

Although the results obtained from deriving word level cues from the candidate list are comparable to the results obtained from direct cue extraction, they are not identical. The two sources of information are also different. It should be possible to increase accuracy by combining both sources of information. For instance, the amount of valid information which can be derived from the candidates generated by the pattern recognizer varies from list to list, and some of the lists are not particularly informative. In such cases, it should be possible to use direct cue extraction to improve overall results. Likewise, direct cue extraction has a higher margin of error than the candidate lists when it comes to estimating word length and so it should be possible to use constraints derived from the latter method in order to limit the values obtained from direct cue extraction.

5.4.2 Method

An experiment was conducted to see whether it was possible to successfully merge the output from direct cue extraction and from the candidate lists.

The modification to the probability scores caused by the presence or absence of the cues ascender and descender are given in Table 5-15. The presence of ascenders or descenders increases the scores of any zones with which they are associated, e.g. if an ascender is present then the middle and upper zone score is increased, and the upper, middle and lower zone score is increased. At the same time, the scores are decrease for any zones with which its existence is in conflict, e.g. if an ascender is present then the middle zone score is decreased because this score represents the probability that data lies solely within the middle zone. The absence of ascenders or descenders has a similar, but opposite, effect on the scores for the zones. The same method that was used in Experiment 5 was then employed to obtain a final result, i.e. the highest score out of each of the zones was used to decide whether the presence of ascenders were indicated, and whether the presence of descenders were indicated. A range of values for increasing and decreasing the probability scores were tested. Detailed results are given in Table K-2 in Appendix K. The selected value was 10.3.

result suggested by candidate list	probability score for the target lying solely within the middle zone	probability score for the target lying within both the middle and the upper zone	probability score for the target lying within both the middle and the lower zone	probability score for the target lying within all three of the zones
ascender present	decrease	increase	no change	increase
ascender absent	increase	decrease	no change	decrease
descender present	decrease	no change	increase	increase
descender absent	increase	no change	decrease	decrease

Table 5-15: Changes made to the probability scores dependent on the results suggested by the candidate list

A simple voting procedure can be used in the case of the cues dot and cross. The number of dots (or crosses) estimated by the word shape recognizer were added to the result suggested by the candidate list (0 if the candidate list suggested that the cue was absent, and 1 if the candidate list suggested that the cue was present). If the resulting number was a 0 then the cue was taken to be absent, otherwise it was taken to present.

It is not possible to combine the word length estimations from the two methods in such a simple fashion because both types of information are the same. There is no way, therefore, of distinguishing in any one instance between the two methods. However, it is possible to derive both a first and second choice for the cues first, last and length (see section 5.2.6) and since, in this case, there are three possibilities it is feasible to make a useful comparison between them. Setting aside those cases where the value suggested by direct cue extraction and the first choice given by the candidate list are identical, Table 5-16 shows how the error rate of the two methods increases according to the difference in length between the values suggested by them. In this table, if the difference in length between the two methods is 5 or over then this means that neither approach is correct.

difference	percent correct using candidate list	percent correct using direct cue extraction	percent neither correct
0	54.0%	54.0%	46.0%
1	33.3%	37.0%	29.6%
2	26.0%	21.0%	53.0%
3	9.5%	28.6%	61.9%
4	0.0%	25.0%	75.0%
5	0.0%	0.0%	100.0%
6	0.0%	0.0%	100.0%

Table 5-16: Percent correct by difference in suggested word length between the two methods: partial data set

5.4.3 Results

Table 5-17 presents a detailed comparison between direct cue extraction and using the candidate list for the partial data set. The last column of Table 5-17 shows percent correct recognition of the target using the merged results. It is possible to merge the two sources of information, in most cases, in order to improve overall accuracy. The only exception to this was the cue dot. In this case, direct cue extraction remains the most accurate of the different methods.

cue	percent correct using the candidate list	percent correct using direct cue extraction	percent both correct	percent candidate list alone correct	percent direct cue extraction alone correct	percent neither correct	percent correct after merging
word length	39.0%	40.5%	21.0%	18.0%	19.4%	41.5%	42.2%
ascender presence/ absence	82.7%	87.8%	74.3%	8.4%	13.5%	3.8%	89.5%
descender presence/ absence	73.3%	93.4%	69.0%	4.3%	24.4%	2.3%	94.2%
dot presence/ absence	79.7%	90.6%	76.8%	3.0%	13.8%	6.4%	80.7%
cross presence/ absence	68.4%	67.2%	47.1%	21.3%	20.1%	11.5%	70.0%

Table 5-17: Comparison between direct cue extraction and using the candidate list, plus merging the two approaches: partial data set

5.4.4 Discussion

This experiment used information extracted from the list of alternatives generated by the recognizer and merged it with information obtained from direct cue extraction. The combination of two different sources of information is a successful approach.

It is also possible to calculate specific error rates for particular combinations between the two sources of information. For example, when the candidate list indicates that crosses are absent and direct cue extraction also indicates that crosses are absent then it is most likely that crosses are absent (error rate of 21.19%). When the candidate list indicates that crosses are present and direct cue extraction indicates that crosses are absent then it is most likely that crosses are present (error rate of 46.44%). Currently such information is not used in the word level method. However, it is possible to exploit the differences between the two sources of information in order to modify the probabilities used by the word level method.

5.5 Conclusions

It has proved possible to improve the way in which cues are derived from the candidate list. An increase in the accuracy of cue detection for all of the cues has been produced. It has also proved possible to develop an alternative source for these cues via direct cue extraction. Direct cue extraction is also important because the candidate list cannot be used for cue derivation in the case of catastrophic failures.

The level of accuracy of the two methods is, in general, similar. Direct cue extraction is, on average, more accurate than using the candidate list for the partial data set. For the complete data set (excluding catastrophic failures) using the candidate list is, on average, more accurate than direct cue extraction. Merging direct cue extraction and the word level method successfully improved the accuracy of detection for the cues length, ascenders, descenders and crosses. The best approach, in practice, is to use a combination of different derivation methods (see section 6.2.3). A final evaluation of the different approaches, using the methods on their own or in combination, is dependent upon the word level method, i.e. the way in which word level cues are to be applied. The particular metric used is not especially important; what is important is that the method chosen minimizes the likelihood of any of the values being incorrect when taken together (see section 5.2.7.1 and Appendix I).

Chapter 6: Improving the Word Level Method

6.1 Introduction

The present chapter investigates the use of imperfect information about word level cues in order to construct a list of word candidates. A method for applying word level cues has been presented in Chapter 4. It is possible to improve upon this method. The central consideration in Chapter 4 was to demonstrate that the word level method could be effective. It was apparent that a number of problems existed with the existing method. Word frequency information was used, but was applied too simplistically. The integration of the pattern recognizer with the word level method should be improved. The training set was a subset of the test set. Other work has shown that it is possible to produce more accurate results than those obtained using the existing method [Bellaby, et. al., 1996a; Bellaby, et. al., 1996b]. The method presented here differs significantly from the approach described in Chapter 4.

The pattern recognizer can be considered to be a conventional character based recognizer which is geared towards giving the target word as the top ranked choice in a set of likely candidates. This is, of course, reasonable since one way to judge output from a handwriting recognition system is on the basis of the target word being top ranked and this calls for one forced choice. However, there is a disadvantage to this approach since the effort to place the target word at the top rank comes at a price. That is that the word lists generated by the pattern recognizer are not well suited for post-processing (e.g. syntactic and semantic analysis) because further selection is not possible unless the target word occurs within the list of alternatives. Post-processing allows further selection to be made from a list of word alternatives and so increase levels of accuracy. For the pattern recognizer to be suitable for post-processing, it is important that the target word is found, regardless of its rank. The pattern recognizer can be considered to be discriminatory (target top ranked) but fragile: it tends either to get the target word correct (to place the target word at the top of the list of alternatives), or to fail to identify the target word at all.

What will be shown is that it is possible to use word level cues to improve the performance of a pattern recognition system. It will be demonstrated that the word level method improves the performance of an even already efficient pattern recognizer. Discrimination and robustness are two different objectives and they do not usually coincide. For example, a pattern recognition system may seek to improve discrimination at the expense of robustness. A pattern recognizer will be unable to recognize a proportion of its input because of the ambiguity of handwriting. A machine recognition system can be considered to be fragile if it completely fails to recognize the target word. The discrimination of a machine recognition system is a measurement of how precisely it manages to identify the target word.

One way to produce a machine system which is both discriminatory (target top ranked) and robust (target found even when it is not top ranked) is to combine different, but complementary recognition methods. The word level method and the pattern recognizer take very different approaches towards handwriting recognition. It is only by integrating different sources of information that a stronger, more robust, machine system can be developed; in particular, such an integrated machine system is needed in order to deal with the particular problems of unconstrained cursive handwriting and a large lexicon.

Section 6.2 through to section 6.5 deal with the development of the word level method. Section 6.2 (Experiment 7) explores a way to apply the word level cues whose derivation was investigated in Chapter 5. Word frequency is the one word level cue relevant to the recognition of cursive script not investigated in Chapter 5 (see section 2.2.6). Section 6.3 (Experiment 8) investigates the use of word frequency information. Section 6.4 (Experiment 9) explores the integration of the word level method and the pattern recognizer. The integration of these two approaches will produce a system which is both discriminatory and robust [Bellaby, et. al., 1996a]. Section 6.5 (Experiment 10) examines a way in which the word level method can be integrated with a bottom-up approach towards recognition. This is an initial attempt to develop a system which allows communication back and forth between different levels of the system. Specifically, this will be a system in which the word and letter levels have an effect on each other [McClelland & Rumelhart, 1981; Rumelhart & McClelland, 1982]. Letter verification procedures have been developed which can be applied to the integrated system. Other work has shown that the use of these letter verification procedures leads to an improvement in machine performance [Bellaby, et. al., 1996b].

Section 6.6 through to section 6.8 present an evaluation of the developed method. It has been stated above that one objective of the word level method is the creation of a system suitable for post-processing. Section 6.6 (Experiment 11) examines whether or not the word level method has met this objective. Experiment 11 examines the potential improvement to be gained by the use of information from the meta-word level. This is a preliminary investigation into the integration of such information into a script recognition system. Section 6.7 (Experiment 13) looks at what makes the word level method different from a word shape recognizer. The recognition performance of the word level method and a word shape recognizer are compared. Lastly, section 6.8 (Experiment 13) explores the relevance of the word level method to the word superiority effect.

The word level method is applied to the input data used in Experiment 1 (section 3.2). It is not claimed that the word level method is the same as the word superiority effect. However, a context effect at the word level has been caused by the word level method. The data and results of Experiment 1 are re-examined to discover what improvement in performance has resulted from the addition of the word level method. The application of the word level method to the input data used in Experiment 1 also means that the word level method will have been applied to another data set, as well as its test data set. This will demonstrate that the word level method is robust.

The full 200 word data set was used as the basis for testing, i.e. all of the words lists irrespective of the ranking of the target word and including uninformative word lists and catastrophic failures. The data set was randomly split into two halves. One half of the original data samples were used to generate the confusion matrices and probabilities whilst the second half were used as test data. Unless otherwise stated, this data set is used for all of the experiments presented in this chapter.

6.2 Experiment 7: Applying Word Level Information

6.2.1 Introduction

The first experiment examines a way to apply word level cues in order to create a list of word candidates. The method presented here improves upon the method presented in Chapter 4. The first step towards improving the word level method is to change the way in which confusions are derived and the way in which probability is used to apply the cues. The word level method is used as a replacement for the pattern recognizer in those cases where the pattern recognizer has experienced a catastrophic failure, that is, where the pattern recognizer has completely failed to generate any output. Word level cues are used in those cases to generate a candidate list in place of the pattern recognizer.

6.2.2 Method

Word level cues are used to derive a new list of candidate words. The result of examining the list of word alternatives, or via pattern recognition, is a pattern of cues. Cue detection is imperfect. Therefore, alternative values for each of the cues are generated. The probability of failure for each cue is calculated by using a training sample and comparing the cues obtained from this sample with their target values. A confusion matrix for each cue is also constructed. Given that cue detection is incorrect, the confusion matrix provides the probability of a given confusion occurring based on the frequency of confusions in the training set.

During testing alternative patterns are then generated using these probabilities. Some of the generated patterns can be rejected, e.g. in those cases where two or more of the cues are contradictory, e.g. "contains the letter 'p'" and "does not contain an descender", or where the generated pattern does not occur in the lexicon. The probabilities used to generate the patterns are used to provide each pattern with a confidence score. Finally, the patterns are sorted according to their score and a lexicon is searched for the words which match each of the generated patterns. In this manner, a list of alternatives was produced for each of the words under examination. The list of words is allowed to grow in size until a predetermined threshold of 100 candidates is reached. This is the same maximum as used by the pattern recognizer. This makes it easier to compare the output of the two different approaches. The threshold of 100 candidates also restricts consideration to the more likely confusion values, but will include instances where the word level method gave the target word as an alternative even though the target word was ranked low (section 4.3.4 and section 4.3.6.2).

A confusion matrix for each of the cues first, last and length is constructed by using a training sample and comparing the cues obtained from this sample with their target values. The confusion matrix contains information about what confusions a particular instance may have, together with the probability of this confusion occurring. For example, in the confusion matrix used for first letter it highly likely that the letter 'a' can be confused with the letter 'c', less likely that it can be confused with the letters 'd' or 'e', extremely unlikely that it can be confused with the letter 'b', and it cannot be confused at all with the letter 'y'.

The success and failure rates for each of the cues are derived. For the purposes of determining probability the cues ascender, descender, dot and cross are considered to one event. This is because they being are treated as statistically dependent (see below). All of the other cues are treated as statistically independent events.

For statistically independent events

$$p(A \ and \ B) = p(A)p(B)$$

Let n_e be the number of confusions for a cue E, n_f the number of confusions for a cue F, n_g the number of confusions for a cue G, and n_h the number of confusions for a cue H.

Let the outcome for each cue be E_i, F_j, G_k, H_l, where E_0 (for example) is the probability of success for the cue E, E_1 is the probability of failure and substitution of the first confusion, E_2 is the probability of failure and substitution of the second confusion, etc.

The probability score for any one pattern being generated from the initial set of word level cues is calculated as follows.

$$\textit{probability score} = p(E_i)p(F_j)p(G_k)p(H_l)$$

Note that

$$\sum_{i=1,\, j=1,\, k=1,\, l=1}^{n_e,\, n_f,\, n_g,\, n_h} p(E_i)p(F_j)p(G_k)p(H_l) = 1$$

Specific instances for all of the cues are used. For example, the cues ascender, descender, dot and cross each have four instances:

1) failure when the cue is indicated to be absent.

2) success when the cue is indicated to be absent.

3) failure when the cue is indicated to be present.

4) success when the cue is indicated to be present.

Specific instances for cues first, last and length are also used, i.e. each character or length has its own associated success and failure rate.

It was decided that using conditional probabilities throughout was not justifiable statistically. The cues ascender, descender, dot and cross are treated as statistically dependent. The other cues are all treated as statistically independent. This decision was made firstly on pragmatic grounds. The number of confusions for letters and word length vary. In practice, there are 209 instances in the confusion matrix for first letter, 123 instances in the confusion matrix for last letter, 85 instances in the confusion matrix for length. In contrast, the cues ascender, descender, dot and cross each have only 4 instances.

The sample size is adequate for these four cues to be combined. This is not the case with the other three cues. Secondly, the cues ascender, descender, dot and cross were chosen to be a set of generalized cues which preserve the sort of information retained across letter confusions (section 2.2.8.1 and section 2.4.2). As such these cues are inter-related in a way that the other cues are not.

The current method assumes that some events are statistically independent. In reality occurrences of all the events are related to the occurrences of other events, and therefore some kind of conditional probability should have been used throughout. It is possible to use conditional probabilities throughout. However, the current data sample is too small to carry this out for all the probabilities used. For example, just under 1,800 word lists are used to test the method. If all of the cues are treated as being related statistically, then a substantial proportion of the correct combinations of cues occur with a frequency of 1. Many apparently valid combinations of cues do not occur at all. Treating the cues first, last and length as statistically independent almost certainly means that the word level method, as it currently stands, is not providing as accurate results as should be the case. The results obtained using conditional probabilities throughout would indeed be very accurate, but the method would not be robust when applied to other data samples. If conditional probabilities are used throughout then all bar two of the target patterns are selected as the first confusion pattern generated by the word level method, and the remaining two are selected as the second confusion pattern generated by the word level method.

It is possible for two or more confusions to have the same score. This is not just a matter of sample size. For instance, the letter 'b' might be equally confused with 'a' or 'c'. This is important because the ordering produced by sort routines is unspecified when the sort criteria are equal. It is likely that groups of words will have the same score, because they are consistent with the same pattern. A second criterion was therefore introduced into the sort routine: *a priori* word frequency (see section 2.2.6 and section 4.4.2). If two words have the same score then the word with the highest word frequency is placed first. However, the ordering produced by the sort routine can still remain unspecified. In order to keep the test results consistent the extra arbitrary criterion of alphabetical ordering was used in the sort routines.

In the case of having both a first and a second letter available, the following modification to the probabilities was used:

1) If a second letter is not available, then continue as normal.

2) If a second letter is available, then look up the probability that the first letter is incorrect.

3) Look up the probability that the second letter is correct.

4) Multiply each probability in the confusion matrix by the probability that the second letter choice is incorrect. The reason for doing this is the probability that a confusion is correct is conditional on the second letter choice being incorrect. The conditional probability of *B*, given that *A* has occurred, is as follows.

$$p(A \mid B) = p(A \text{ and } B) \,/\, p(A)$$

5) If the second letter exists in the confusion matrix, add together the probability that the second letter is correct and the new probability for that letter in the confusion matrix.

In the case of catastrophic failures (i.e. where the pattern recognizer had generated no candidates at all) word level cues were used to generate new candidates in the same way as above. However, a larger number of possible candidates than normal were allowed; candidate lists containing 300 word alternatives instead of 100 words were generated. This means that word frequency information will be allowed to have a greater influence on the final ordering of the candidate list than in the case of non-catastrophic failures.

6.2.3 Results

The word level method was more effective when a single choice for the cues first letter, last letter and word length was used, than when multiple choices for these cues was used (see section 5.2.6). The reason for this is that the probability that the first confusion is correct when a single choice is used, is greater than the probability that the second choice is correct when two choices are used (see section 5.2.7.5). The accuracy of cue detection for the first choice when two choices are derived is also slightly lower than the accuracy of cue detection when only a single choice is derived.

Merging direct cue extraction and the word level method successfully improved the accuracy of detection of individual cues (see section 5.4). However, when these cues were used in conjunction no improvement to the word level method resulted. A combination of different derivation methods has proven to be the best approach (see section 5.2.7) and Appendix I). Using the candidate list to derive some of the cues and direct cue extraction to derive the rest of the cues is more effective than using either independently. The best combination of methods minimizes the likelihood of any of the values being incorrect when taken together. In this experiment, some cue values were estimated from the candidate lists, and some from direct pattern recognition, to give the combination that produced the best results. Naturally, only direct cue extraction could be used in the case of catastrophic failures.

The method of applying word level cues described in this experiment is an improvement on the method presented in Chapter 4. Table 6-1 shows the results of the current word level method with the method presented in Chapter 4. In Chapter 4, results were given for the word level method in those cases where the pattern recognizer had generated some candidates but had completely failed to identify the target word (see section 4.4.3 and Table 4-10). Results for the current word level method are provided for identical cases in the test data set. Results for the current word level method compared to the method presented in Chapter 4 show an increase in all three rows.

	word level method described in Chapter 4	word level method described in Chapter 6
top ranked	8.4%	222.3%
top 5	26.3%	38.1%
anywhere in list	50.1%	71.6%

Table 6-1: The word level method described in Chapter 4 vs. the word level method described in Chapter 6. Percent correct recognition when the pattern recognizer failed to provide that target word as an alternative.

Figure 6-1 shows the percent correct recognition for

- the pattern recognizer (PR)
- the word level method (IWLM)

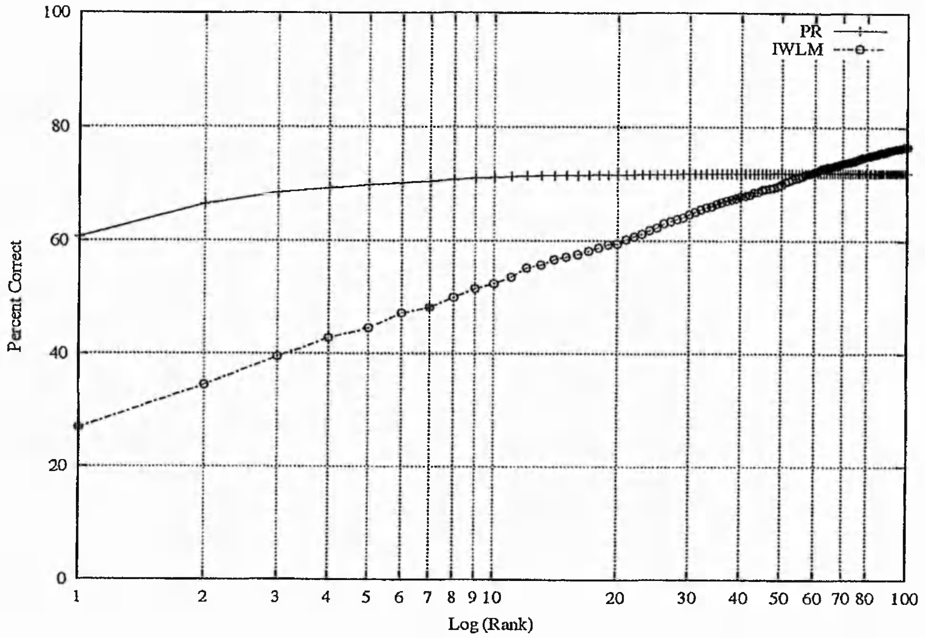


Figure 6-1: The word level method vs. the pattern recognizer. Percent target word correctly identified at, or above, that rank

The average number of candidates generated by the word level method was 93. In the test set the average number of candidates generated by the pattern recognizer was 14.

6.2.4 Discussion

This experiment demonstrates an improvement in the word level method compared the method described in Chapter 4. It has proved possible to derive usable information about word level cues from the list of word alternatives given by the existing pattern recognition software. It has also proved possible to derive the same kind of information using pattern recognition and, subsequently, to apply the information from these two separate sources. Finally, it has proved possible to use word level cues to successfully derive the target word in a significant number of cases.

A threshold of 100 candidates has been imposed to restrict consideration to the more likely confusion values. If all of the confusion matrices included each and every possible confusion (no matter how unlikely), and no threshold was imposed, then the word level method would generate the complete lexicon every time it was applied. The confusion matrices that have been used in this experiment do not include each and every possible confusion because of the restrictions of the data sample used to calculate the confusions, e.g. none of the data samples begins with the letter 'x' and so 'x' does not appear in the confusion matrix for first letter. However, the confusion matrices do include the target value for the overwhelming majority of the target words. If a larger threshold was used then the number of target words unrecognized would be smaller.

The general effectiveness of the cues can be demonstrated by them cues to re-order the original list of word alternatives generated by the pattern recognizer. Word level cues were used to re-order the word list generated by the pattern recognizer. A confidence value was ascribed to each cue that has been identified. If a candidate matched the identified cue then the confidence score given to the candidate was increased appropriately. The list of words was then ranked on the new confidence scores. *A priori* frequency was not used to order the candidates from the pattern recognizer.

Figure 6-2 shows the percent correct recognition after the word level cues were used to re-order the candidate lists generated by the pattern recognizer. Results are given for

- the pattern recognizer re-ordered using word level cues (PR re-ordered)

- the pattern recognizer (PR)

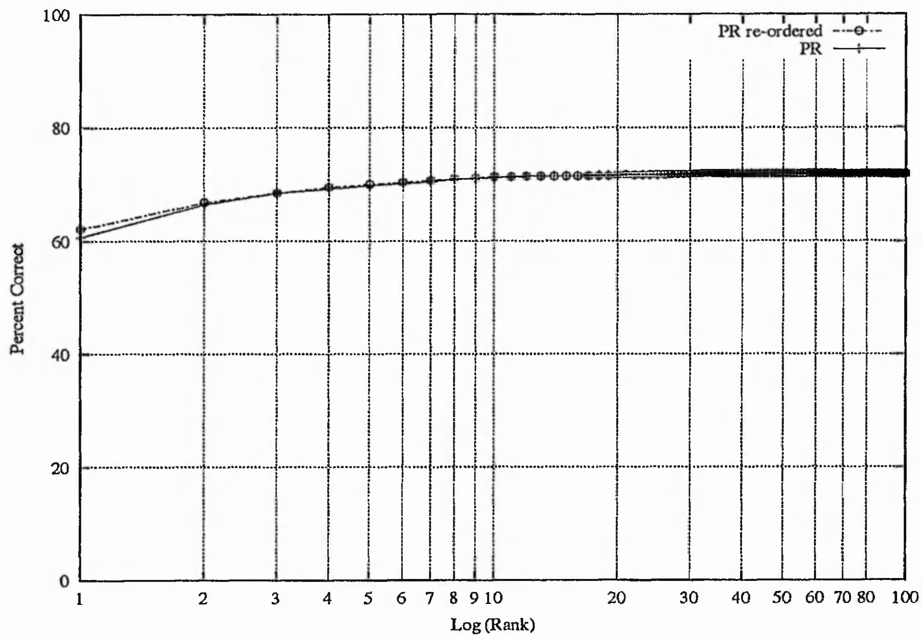


Figure 6-2: The pattern recognizer on its own and re-ordered by the word level method. Percent target word correctly identified at, or above, that rank

6.3 Experiment 8: Word Frequency

6.3.1 Introduction

The frequency of occurrence of words is not random. Certain words occur more frequently than others in written text and the word probability can be calculated by counting the number of instances of the word in a corpus. Even without a surrounding context domain information could be used to bias the result towards particular domain specific words.

Word frequency was applied too simplistically in Chapter 4. A simple bias towards high frequency words will swamp low frequency words. A balance between the scores generated by the word level method and word frequency information needs to be found [Srihari & Bozinovic, 1982; Hull, et. al., 1983; Srihari, et. al., 1983].

6.3.2 Method

Word frequency can be shown to make a contribution to the Word Superiority Effect. An experiment (experiment 8) was conducted into the effect which word frequency information would have on the candidate lists generated by the word level method. The effect of word frequency would be determined empirically. The reason for this is that the score given by the word level method to a candidate is unrelated to its *a priori* frequency score.

The score ascribed to each candidate by the word level method is only a measure of the confidence that a candidate matches the target word relative to the confidence that other candidates in the same list of words also match the target word (see section 6.2).

The probabilities derived and used to generate the new list of candidates are not an accurate representation of the absolute probability that a candidate is correct, nor of the relative probabilities that two different candidates are correct, e.g. C1 has a score of 0.2, and C2 has a score of 0.1. It can be stated that the word level method is more confident that C1 is correct than C2. However, it is not possible to state that C1 is twice as likely to be correct than C2.

The reason for this is that the probabilities used to generate the candidates during the application of the word level method do not always generate a result, e.g.

they can be associated with a an empty group (i.e. the pattern does not correspond to any of the words in common lexicon).

For example, the highest probability is associated with the original shape, but the original shape may not correspond to any words in the lexicon, and may indeed be completely invalid, e.g. ascender absent but the first letter is a ‘d’.

As an initial test, the lists were sorted on the basis of word frequency, with their word level method score only as a secondary criterion. This had the effect of decreasing the number of target words which were highly ranked, but also promoted a number of words which were originally ranked low. This test also demonstrated that the ordering of the final lists is not completely determined by word frequency since the lists differed considerably. The effect of word frequency was thus shown to be useful but, without some sort of restraint, too large.

The effect of word frequency was dampened so that the final ranking of the candidates was influenced to a greater extent by the score associated with each word than it was by word frequency. This was done by dividing word frequency by a constant and then taking the square root. This has the effect of making the distribution flatter.

A range of constants was tested. The results of this test are given in Table L-1 in Appendix L. A value of 100 was shown to be effective and was used to produce the results given in Figure 6-3. If dividing the word frequency by 100 resulted in a score less than 1, a score of 1 was used. The score given to each candidate by the word level method was multiplied by the value calculated from the word frequency to give a final score. The list of alternatives was then ranked according to their final score.

6.3.3 Results

Figure 6-3 shows the percent correct recognition for the word level method after the application of word frequency information. Results are shown for

- the word level method after the application of word frequency information (WLM)
- the word level method without word frequency information (IWLM).

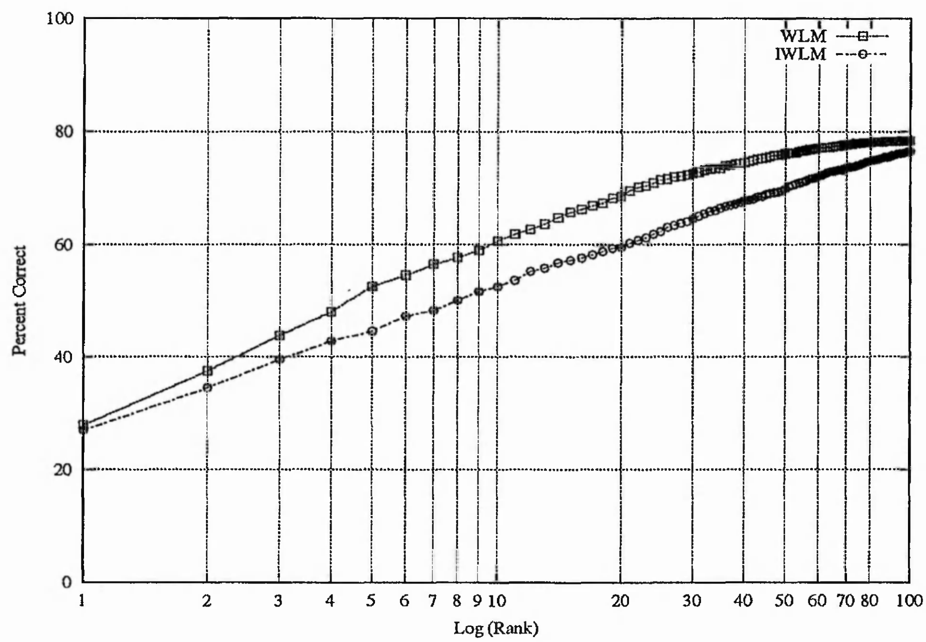


Figure 6-3: The word level method combined with frequency information.
Percent target word correctly identified at, or above, that rank

6.3.4 Discussion

The idea that the lexicon should be ordered on the basis of word frequency is not new [e.g. Forster, 1976; Wilson, 1984]. The approach taken in this present work attempts to relate word frequency both to lexical information and word level information, using machine based methods to extract this latter source of information. The introduction of word frequency into the recognition process raises an important issue; the system must still be able to recognize low frequency words when they occur. The approach which has been taken towards the integration of word level information directly addresses this issue.

6.4 Experiment 9: Integration

6.4.1 Introduction

The original version of the word level method that was presented in Chapter 4 merged the top 3 candidates from the list of alternatives generated by the pattern recognizer with the candidates produced by the word level method Since the version of the word level method presented in this chapter is more accurate than the version described in Chapter 4, such an approach can be improved upon. The outputs of the pattern recognizer and the word level method are merged using a similar method to that described in Chapter 4 (see section 4.4.2). The word level method, like the pattern recognizer, produces a ranked candidate list and it is possible to ascribe a score to each of the candidates and thus merge the output of the two different methods. Previous experiments have shown that the word level method is not as discriminatory as the pattern recognizer. However, they have also demonstrated that the word level method produces significantly different results to that of the pattern recognizer, and they have indicated that the word level method is less fragile than the pattern recognizer. It is possible to combine the pattern recognizer and the word level method in order to improve machine performance.

6.4.2 Method

The pattern recognizer generates a list of word alternatives, each of which has an associated confidence score. This score is in the range 1 to 100. In order to merge the list of word alternatives generated by the pattern recognizer with the list of words generated by the word level method the latter had to be given a comparable value.

A score has been given to each of the alternatives generated by the word level method (see section 6.3). This score is based on the probability of the pattern's occurrence and the word frequency of the alternatives. This score is used to rank the word alternatives.

A new dummy confidence score was given to the words in the candidate list generated by the word level method. An initial dummy value was given to the top ranked candidate and each subsequent word in the list was given a score one less than its immediate successor, e.g. the top ranked word was given a score of 37, the second a score of 36, and so on. A range of values for the dummy score given to the candidates generated by the word level method were tested. Detailed results are given in Appendix M. A dummy score of 37 (out of 100) proved to be the most accurate value. The value of 37 had been chosen empirically by comparing the accuracy of the pattern recognition system with the accuracy of the word level method.

Nothing would be gained by using a more complex scoring method. The score given to the candidates produced by the word level method is unrelated to the confidence score given to the candidates produced by the pattern recognizer. There is, therefore, no meaningful measure of accuracy relating the two lists of words (see section 6.3.2).

The list of alternatives generated by the pattern recognizer was merged with the alternatives produced by the word level method. The resulting merged list was ordered on the basis of the confidence score given to each word alternative.

6.4.3 Results

Figure 6-4 shows the percent correct recognition after merging the word level method with the pattern recognizer. Unless otherwise stated all results for the word level method given here, and below, include the effect of word frequency. Results are given for

- the word level method merged with the pattern recognizer (PR+WLM)
- the pattern recognizer (PR)
- the word level method plus word frequency (WLM)

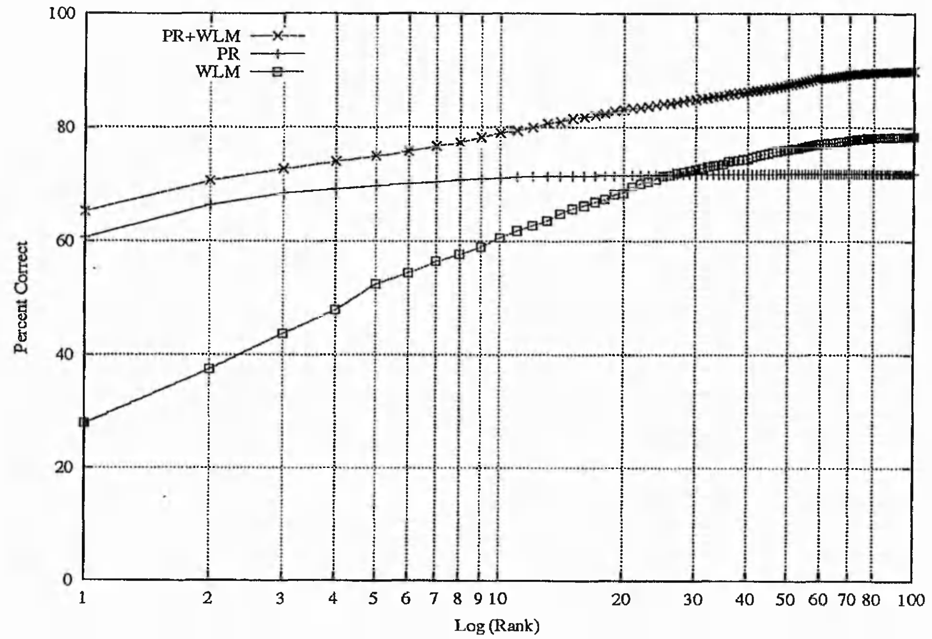


Figure 6-4: The word level method merged with the pattern recognizer. Percent target word correctly identified at, or above, that rank

A significant decrease in the error rate has been produced. The failure rate has dropped from 28% to 10% failure rate after the application of word level information. A small increase in the best rate has also been produced.

6.4.4 Discussion

It has been demonstrated that it is possible to successfully merge the outputs of the pattern recognizer and the word level method. The word level method is not intended to be an alternative to the pattern recognizer. A specific point in the recognition process for the application of this approach has been identified. It is argued that pattern and word level recognition play distinct but complementary roles in the machine recognition process. The consequence of integrating these different approaches is an overall improvement in machine performance. The word level method is more robust, but less discriminatory, than the pattern recognizer; it tends to find the intended word but has problems ensuring that the intended word appears towards the top of the list. It has been demonstrated that it is possible to successfully merge the outputs of the pattern recognizer and the word level method. Two recognition methods which display significantly different characteristics can be integrated to improve machine performance. The consequence of integrating these two methods is an increase in discrimination and robustness. In other words, the proportion of target words and those top ranked are both increased. A handwriting recognition system which is both discriminatory and robust has been therefore been created. It is therefore argued that the word level method is a viable way to cope with ambiguous or incorrect output from a pattern recognizer. These results are very encouraging. The experiment has shown that useful information can be gained even when the recognizer did not identify the target word. Any positive result in the case of such recognition failures would be significant, but the word level method has actually proved very successful.

The results showed an increase in the proportion of target words top ranked (from 61% to 65%), a larger increase in the proportion of target words in the top 10 (from 71% to 79%) and an overall increase in the proportion of targets found in the top 10 (from 72% to 90%).

Since ceiling effects will minimise any possible improvements, these figures are significant. Just considering those cases where the recognizer failed to produce the correct word, the word level method derives the word more than half of the time. Other methods of merging the lists could further increase the proportion of target words top ranked: the use of other knowledge sources could also bring improvement. The fact that the word level method increased the proportion of targets found increases the chances of such improvement.

Some elements of the information which are used by the word level method can also be used by the pattern recognizer. However, this information is being used in two completely different ways. The word level method re-organizes this information for its own purposes and structures this information in a different fashion to that of the pattern recognizer. As stated in Chapter 1 the data used throughout this thesis was generated using the strict pattern recognizer. Alternative pattern recognizers have also been developed by Powalka. The new pattern recognition methods are more accurate at identifying the target word and also have a lower error rate. This helps the overall accuracy of the word level method.

It has been shown previously that training can decrease a pattern recognizer's ability to unambiguously identify cursive script. The indications are that further training will improve the accuracy of word level cue extraction. However, it is not known what limitations to accuracy may exist. It may be the case that extensive training will have a deleterious effect upon accuracy. However, the word level approach uses information derived from a list of word alternatives. This means that the disadvantages caused by training which were described in Chapter 3 will have less impact upon the accuracy of word level cues than they do upon pattern recognition. Training will result in certain improvement over the short term, but its effect over the long term will be an increase in ambiguity.

It is possible to identify a word using word level cues regardless of what actual method of recognition is used. The recognition method is irrelevant, except in terms of its accuracy, to the employment of word level cues to identify a word. It is not just a case of adding these cues to a machine system, indeed some systems may already utilize one or more of the cues indicated. Rather these cues can be used either alone, or in conjunction with other contextual information, to successfully identify many words. Attempts have been made previously to develop a two-stage recognition system [e.g. Hull, et. al., 1983]. Additions to the recognizer developed at the Nottingham Trent University, such as zoning information and other types of feature detection, also implement certain types of word level cue extraction. However, the use of word level information is not a mechanical addition to existing recognition software. It is not envisioned, for example, that additions to Nottingham Trent's pattern recognition system will, or can, supersede the word level approach. These cues have, to some extent, been derived from the psychological literature and it has been demonstrated here that they can be effective without recourse to further pattern recognition.

Although word level cues are currently being used to supplement the operation of the pattern recognition system in this manner, it should be realised that the success of applying word level cues to generate a new list of candidate word lends some support to the argument that it is possible to use a word level approach as the major method of word recognition. The occurrence of catastrophic failures is a case in point. The successful application of the word level method in the context of catastrophic failures suggests that the word level method could be used as the initial stage of a top down approach. The word level method generates the target word when the pattern recognizer cannot since the pattern recognizer must correctly recognize each and every one of the characters in the target word. The word level method does not have this restriction. However, word level cues are often sufficient to find the target word even using imperfect information [Hull & Srihari, 1986; Ho, et. al, 1992a]. The proportion of target words top ranked by the word level method in the case of catastrophic failures was 8%, whilst the proportion of target words placed in the top 100 by the word level method was 72%.

The list of words generated by the word level method will contain candidates not proposed by the pattern recognizer but which have a greater likelihood of resembling the target word than the candidates proposed by the recognizer. The merging of the word level method with the pattern recognizer has therefore made the machine system more robust.

For some applications, it is more acceptable for a handwriting recognition system to present the user with some sort of response than with no response whatsoever, even when the target word does not appear in the list of alternatives generated by the system. The word level method uses similar cues to the human reader. This therefore should mean that recognition mistakes, when they occur, will reinforce appropriate cues as to how the writer modifies his or her writing.

The word level method allows the use of frequency information, there is therefore a greater likelihood statistically of recognizing the target given ill-formed input than using a pattern recognition system which cannot use frequency information. It is more useful for certain applications if a machine system is more accurate in its recognition of words which occur frequently than if the accuracy of a machine system is frequency independent, even if accuracy on average is higher in the latter case.

It can also be argued that, in some circumstances, a user would be more forgiving of a machine system which misrecognized low frequency words than high frequency words since low frequency words tend to be seen as more 'difficult' than high frequency words and, because of the word frequency effect, the reader will tend to make such a pattern of mistakes him- or herself.

There are several reasons why the word level method works. The word level method can relax some of the requirements of the pattern recognizer (e.g. identification of each letter), it abstracts from the information to operate in a radically different fashion from the pattern recognizer, and it uses information which has nothing to do with the pattern of strokes (e.g. statistical information about the lexicon, typical confusions, etc.). The word level method also imposes constraints which the pattern recognizer does not (e.g. word frequency).

It is possible to use pattern recognition to recognize the cues which are used by the word level method. It has proven useful to do so. However, the derivation of word level cues from a candidate list is a method which can be implemented without the need for a new pattern recognizer. The word level method can therefore be integrated using any existing pattern recognizer which generates a list of alternatives. It is therefore possible to introduce the word level method quickly into a machine system. Whilst the results of the present work suggest that a feature recognizer should improve the accuracy of the word level method, the machine system will be improved even if this is not done.

The information derived from a pattern recognizer and the information derived from the candidate list are not identical. The two sources of information overlap, but they are not identical. This means that the candidate list is a new source of information which can be used in conjunction with direct cue extraction.

The conclusions presented here are applicable to other machine recognition systems. The word level method is generating an appropriate response to the input data. A strong correspondence exists between the results of the pattern recognizer and the word level method when it is used as the sole recognition method. This implies that the word level method is describing valid information which is not merely tied to the particular data sample which has been used

The data sample used represented a variety of different writing styles, and it also varied in the quality of the writing (or, to be more specific, the accuracy of the pattern recognizer on the different writers varied).

The letter confusion matrices which have been generated and used in this work appear robust (see Appendix N). The confusion matrices bear a partial resemblance to the confusion matrix observed by Bouma [Bouma, 1971]. However, Bouma used printed letters in his experiment and double (or triple) letter confusions were not examined. The confusion matrix generated matched expected confusions. However, further experimentation is needed before a proper evaluation of the confusion matrix and a comparison between the confusion matrix used by the machine system and the typical confusions made by human readers could be made.

Any confusion matrix will bear some similarity with the one used by the word level method, e.g. it would have to be a very unusual recognizer that confused letters such as 'b' and 'g', or 'p' and 'c'. Likewise length will bear approximate resemblance between target and confusion, e.g. one would not expect a word of length 12 to give rise to a word length 1. It should be noted that the confusion matrix also reflects the statistical characteristics of the lexicon. The confusions are therefore not solely a consequence of misrecognitions by the pattern recognizer.

However, further work is needed using a larger data sample to complete this work. For example, the sample size was not large enough to generate reliable probabilities for confusions of the letter 'x'.

6.5 Experiment 10: Letter Verification

6.5.1 Introduction

An important consequence of the word level method is it allows a fully integrated system to be developed, e.g. a system which allows communication back and forth between different levels of the system. The word level method can be considered a higher level process which adds new candidates. It is therefore possible to re-examine the physical evidence in the light of the new candidates which have been proposed by the word level method. The word level method has no way of distinguishing between candidates delineated by the same seven word level cues except by their word frequency. However, it is possible to use pattern recognition methods to make a further selection.

6.5.2 Method

The list of words generated by the pattern recognizer was merged with the alternatives produced by the word level method using the method described above (see section 6.4). The resulting merged list was ordered on the basis of the various confidence scores given to each word alternative. The merged word lists were presented to a further letter verification recognizer which applied letter verification procedures on the handwriting samples. Candidates were increased in rank on the basis of the proportion of their characters which the letter verification recognizer recognized within the sample.

The letter verification methods used in this experiment were developed for the word shape, recognizer [Powalka, et. al., 1994; Powalka, 1995]. However, the letter verification methods can be used in isolation as a stand-alone letter verification recognizer. The list of candidates is presented to the letter verification recognizer. Letter alternatives are located and recognized by the recognizer and subsequently matched against the letters of which a candidate is composed. A graph of letters is produced by the letter verification process. Each letter alternative is given a confidence score. The confidence scores of the recognized letters are summed together. This total is divided by the number of letters in the word. Some letters will not have scores. This means that words with more recognized letters will have better scores than those with fewer recognized letters. For example, the score for a single character word whose character has been identified by the letter verification recognizer will be greater than the score for a multiple character word which only had one character identified by the letter verification recognizer.

New scores are calculated using the scores provided by the letter verification recognizer and the merged pattern recognizer and word level method. Scores are averaged together. In this way, the confidence of letters from both sources are boosted with respect to letters proposed by only one source. The new score is used to rank the word alternatives. A weighting factor is used to take account of the relative confidence of the methods. The score obtained from the letter verification recognizer was first multiplied by a constant. A range of constants were tested. The results of this test are given in Table O-1 in Appendix O. A value of 0.76 was shown to be effective and was used to produce the results given in Figure 6-5.

6.5.3 Results

Figure 6-5 shows the percent correct recognition after applying letter verification. Results are shown for the individual methods alone and in combination:

- letter verification applied to the word level method merged with the pattern recognizer (PR+WLM+LV)

- the word level method merged with the pattern recognizer (PR+WLM)

- the pattern recognizer (PR)

- letter verification applied to the word level method (WLM+LV)

- the word level method (WLM)

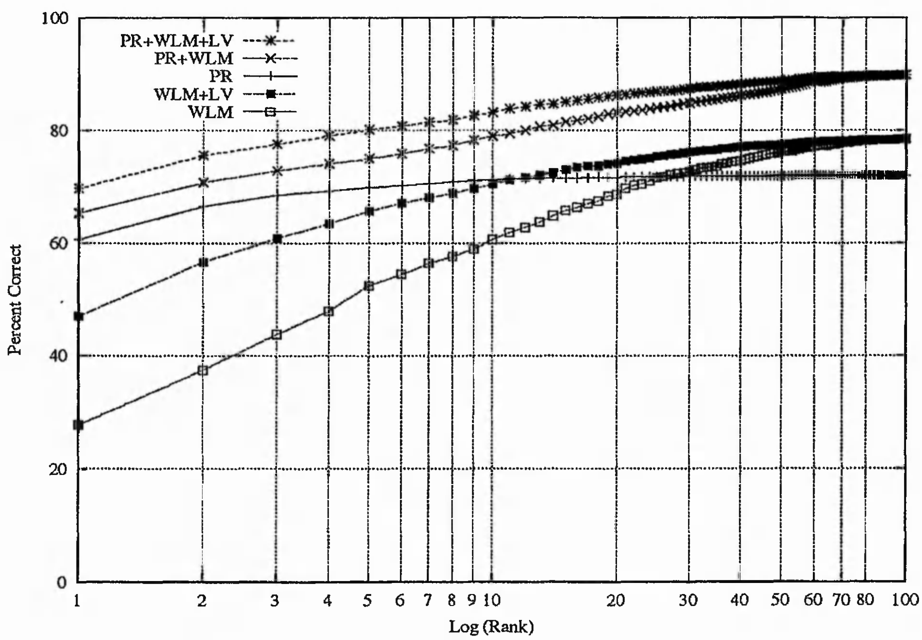


Figure 6-5: Recognition results for the word level method and the pattern recognizer on their own and in combination. Percent target word correctly identified at, or above, that rank

6.5.4 Discussion

Letter verification had a greater impact upon those words where the pattern recognizer suffered a catastrophic failure. This is because the word level method is the sole method used in the case of catastrophic failures. The word level method lacks the ability to discriminate between words selected using the same set of cues. Letter verification adds such a fine-grained ability to distinguish between similar candidates.

Letter verification improves the accuracy of the integrated recognition system. Letter verification does not rely on correctly recognizing each and every letter in the same way as the pattern recognizer. Letter verification provides supporting evidence for those candidates which have been generated by the word level method. However, targets correctly recognized by the pattern recognizer should be corroborated during letter verification. Machine performance on such words should therefore not be degraded.

Results for merging output from the pattern recognizer and the word level method show an increase in all three rows. Letter verification caused a further increase in the proportion of target words top ranked and the proportion of target words in the top 10. The performance of the pattern recognizer on some of the writers was already very accurate. The word level method will have little impact upon recognition performance for such writers and this therefore limits the amount of improvement possible.

The combined method was unable to place the target word in the top 100 candidates 10% of the time. A list length of 100 was used here as it seemed a reasonable length to evaluate the potential effectiveness of the methods and combinations under investigation. A greater proportion of the target words would be found if a larger number of word alternatives were considered. However, the longer the list, the less meaningful it becomes and the greater the computational demands that will be made. The length of the list to be generated must be determined by considerations particular to any application or to any research aims, such as accuracy requirements, post-processing techniques to be used, computational complexity, etc.

6.6 Experiment 11: Higher Level Context

6.6.1 Introduction

It has proven possible to use word level cues for isolated words. If a word is embedded in text then it also becomes possible to use sources of information taken from the meta-word level. Higher level contextual information can be used to further augment performance. Methods for implementing syntactic and semantic information to aid script recognition have been reported [Keenan & Evett, 1994; Lesk, 1987; Liddy & Yu, 1994; Madhvanath & Govindaraju, 1998; Malburg, 1997; Oh, et. al., 1995; Rose & Evett, 1993; Srihari & Baltus, 1993]. The present experiment was carried out as a preliminary investigation into the integration of such information into a script recognition system using the current approach.

A preliminary examination of the data produced by the recognition system suggests, firstly, that the use of syntactic information will be effective and, secondly, that in most cases it is possible to employ very broad syntactic classes, such as tense (e.g. "walk"/ "walking"/"walked"), or number feature (e.g. the distinction between singular and plural: "the boys is"/"are hungry"), in order to produce a strong effect

It is therefore not necessary to introduce a sophisticated parser capable of dealing with more complex or subtle syntactic classes. For example, the following output was obtained from the recognition system:

responsibility

responsibilities

It is not uncommon for the recognition system to produce a list of alternatives that contain morphological variants of one word. The use of broad syntactic information in cases such as this will be of obvious benefit. The system currently in use only distinguishes 47 separate syntactic classes.

6.6.2 Method

The word lists generated in Experiment 9 were used (see section 6.5). The effect of syntactic analysis on these word lists was simulated. In those cases where two or more alternatives were indicated further selection using syntactic class was possible. Each of the candidate lists was re-ordered on the basis of syntactic class. The syntax analyser has about a 97% accuracy in correctly tagging words [Keenan, 1993]. For 97% of the lists (chosen at random) those word alternatives which were in the same syntactic class or classes as the target word were simply placed above those which were dissimilar. The position of words relative to other words of the same syntactic class was maintained. For the remaining lists those word alternatives which were not in the same syntactic class or classes as the target word were placed above those which were. This carried out 100 times and an average of the results calculated.

6.6.3 Results

Figure 6-6 shows the percent correct recognition after the application of syntactic information. Results are given for

- the merged lists sorted by syntactic class (SYN)
- letter verification applied to the word level method merged with the pattern recognizer (PR+WLM+LV)
- the pattern recognizer (PR)
- the word level method (WLM)

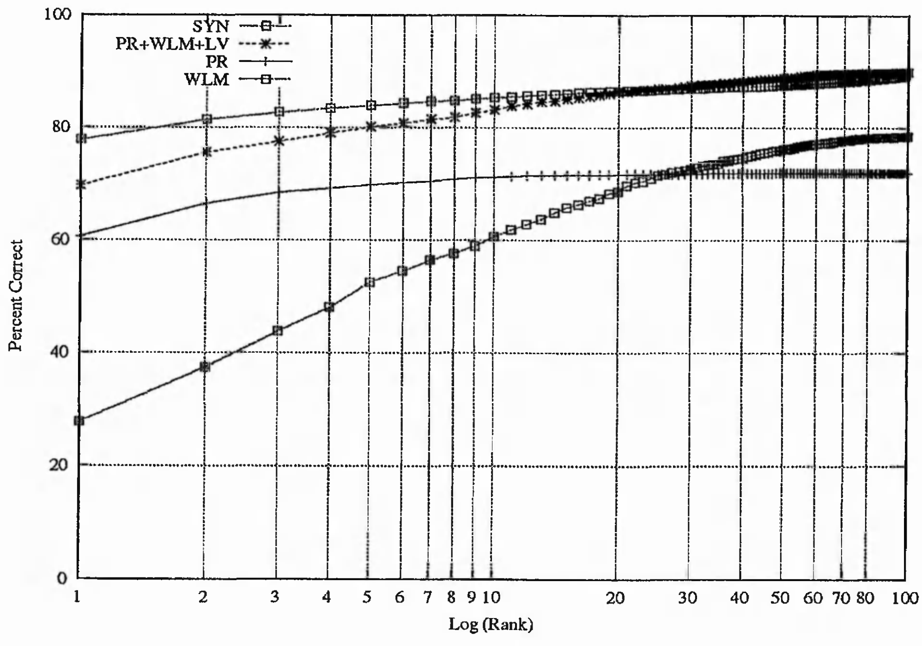


Figure 6-6: The word level method both with and without context. Percent target word correctly identified at, or above, that rank

This experiment indicates the relevance of syntactic analysis. The accuracy of the syntax analyser must be high enough to cause an improvement in recognition performance. The characteristics of the selected criteria (in this case syntactic class) must be appropriate, else they would have only a minor impact on the results. If further improvement was not likely, then it would be more useful to explore other avenues. Current work is concerned with integrating other contextual knowledge sources (see section 7.2).

6.6.4 Discussion

The previous experiments have shown that in a significant number of cases word level cues can be used to derive the target word. This experiment demonstrates that syntactic information can be used successfully to select the target word from a list of alternatives. The use of both of these sources of information in conjunction places the target word at the top of the list of alternatives in 78% of the cases, and in the top 5 in 85% of the cases. Keenan reports that the proportion of target words assigned top rank by the syntax analyser was between 70% and 87%, depending on the test data [Keenan, 1993].

Syntactic information has been used in a very simplistic fashion; word candidates in the same syntactic class or classes as the target word were placed above those which were not. Many words belong to more than one syntactic class. Some words show a high degree of syntactic ambiguity. For example, the word "western" can be a proper noun or an adjective, and the likelihood of the word being used in either syntactic form is approximately the same. However, this is not the case with many words. One syntactic form of the word tends to have a higher frequency of occurrence. For instance, the word "ability" can be used as a singular noun or as a proper noun, but it is far more likely to be used as a singular noun than it is as a proper noun. The word "ability" appears 4,839 times in the Oxford Corpus. In 99.79% of these cases it is used as a singular noun. It is only used as a proper noun in 0.21% of these cases. The probability of occurrence of particular syntactic classes is taken into account by the syntactic analyser developed by Keenan [Keenan & Evett, 1994].

As work continues the results shown in Figure 6-6 will be expanded to include the effects of introducing semantic information, both on its own and in conjunction with syntactic information. Syntax has only been applied in a very simplistic manner here. However, it is possible to use context to determine the probability of syntactic class for any word position.

It is also apparent that the use of semantic information will be useful. A second example taken from the recognition system will serve to illustrate the benefit of semantic information. Given a substitution set:

right

eight

fight

fright

The target word is "eight" and it occurs within the sentence "an octopus has eight arms."

In this particular case, the recognition software has had difficulty in deciphering the initial letter of the target word. Semantic information can be used in this instance to select the more likely of these alternatives in the given context, e.g. dictionary definitions [Rose & Evett, 1992]. Srihari & Baltus report that an increase in top choice word recognition rate from 80% to 95% is possible when context is used [Srihari & Baltus, 1993].

6.7 Experiment 13: Comparison with Word Shape Recognizer

6.7.1 Introduction

The use of word shape information in pattern recognition systems has been suggested by other writers. Examples of recognizers which use word shape are given in section 2.3. It can be argued that a word shape recognizer, like the word level method, attempts to use word level information, albeit implicitly. However, the word level method is not a word shape recognizer. The differences between the two approaches will be discussed below.

A word shape recognizer has been developed [Powalka, et. al., 1994; Powalka, 1995]. The shape of the word is described using the number and location of vertical bars, the height of the vertical bars, zoning information and the number of potential letter positions. Word shape information is matched against a database of stored shapes. A letter verification procedure is used to select from the word alternatives that correspond to a word shape

A comparison will be made between the recognition performance of the pattern recognizer, the word shape recognizer and the word level method.

6.7.2 Method

An experiment was conducted using the word shape recognizer. The handwriting samples and lexicon that were used to test the word level method were once again used (i.e. the 200 word data set and the common lexicon).

The word shape recognizer produces different results to that of the pattern recognizer. The word shape recognizer, like the pattern recognizer, produces ranked candidate lists with each of the word alternatives having an associated confidence score. However, the range and characteristics of the confidence scores are different to the pattern recognizer. The performance of the word shape recognizer was recorded. The output of the word shape recognizer was then merged with the output of the pattern recognizer. Details are given elsewhere [Powalka, 1995]. In brief, a word alternative proposed by both recognizers has its score increased proportional to the higher of the two scores. The combination of results is also biased towards the pattern recognizer.

6.7.3 Results

Recognition performance for each method alone, and some combinations, are given in Figure 6-7. Results are shown for

- letter verification applied to the word level method merged with the pattern recognizer (PR+WLM+LV)
- the word shape recognizer combined with the pattern recognizer (WSR+PR)
- the pattern recognizer (PR)

- the word shape recognizer (WSR)
- the word level method (WLM)

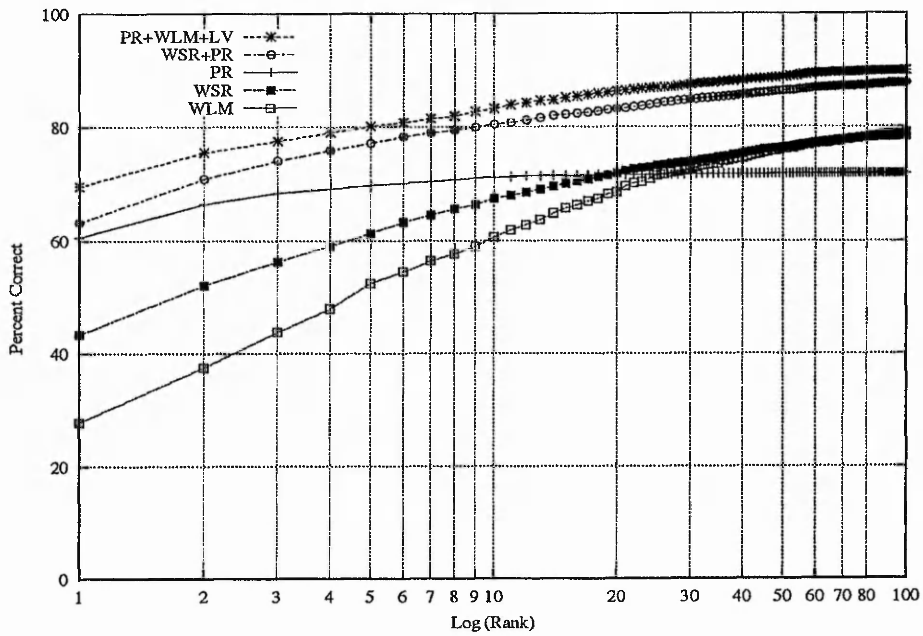


Figure 6-7: Recognition results for the methods on their own and in combination. Percent target word correctly identified at, or above, that rank

The word shape recognizer is less fragile than the pattern recognizer, but lacks its fine-grained recognition abilities. Letter verification applied to the word level method merged with the pattern recognizer is more accurate overall than the word shape recognizer, on its own or combined with the pattern recognizer. The word shape recognizer includes letter verification, so this difference in performance must be caused by the word level method.

6.7.4 Discussion

Different sources of information can be drawn upon for handwriting recognition. Three sources of information have been considered: character-segmentation information, word shape information and lexical information. The methods used to extract these three sources of information are, respectively, a traditional pattern recognizer, a whole word recognizer and a method which uses word level cues. It may be argued that pattern, whole word and word level recognition play distinct but complementary roles in the machine recognition process.

The word level method is not a word shape recognizer. The word level method takes a different approach to handwriting recognition from that of a word shape recognizer. The choice of cues used by the word level method has been influenced by work in Cognitive Psychology. For example, the cues used by the word level method are an attempt to generalize from the way in which letter shapes are retained, by human readers, across confusions. The word level method has created a context effect at the word level. This context effect is similar, if not the same as, the Word Superiority Effect.

The word level method takes a very different approach to handwriting recognition from that of the word shape recognizer. The word level method works on the basis of lexical information, not on the physical evidence. The word level method is a top-down method which uses abstract word level cues. The word level method uses these cues to generate new candidates and subsequently a bottom-up method (letter verification) is applied to these candidates. The word level method moves from lexical information down to word candidates, and from the word level down to the letter level.

Relevant data from the word shape recognizer has been used as the basis for direct cue extraction. Information from the word shape recognizer has been used, after processing, to derive the cues word length, ascenders, descenders, dots and crosses. The pattern recognizer has been used to derive the cues first letter and last letter. Whilst it can be argued that some of the information which the word level method exploits is the same, or similar, to that of a word shape recognizer, it exploits the information in an entirely different way. The interpretation and subsequent use of this information is distinctive. The word level method displays different characteristics to that of a word shape recognizer because it uses different sources of information and takes a very different approach towards recognition. The word shape recognizer attempts to recognize physical features and subsequently match them against a known database of word shapes, e.g. it attempts to match the size and vertical position of the encoding bars. The word level method abstracts from the data generated by the recognizer, e.g. it is unconcerned with the actual number or positioning of ascenders and descenders.

Word shape information is not the same as word level cues. The word shape recognizer relaxes some of the constraints of character and segmentation recognition in favour of word shape recognition and subsequent character verification. However, word shape is still a representation of the physical properties of the input word, e.g. number and position of vertical bars in the input data. The pattern corresponds to the physical form. In contrast, the cues used by the word level method are not simply physical features, but are abstract, i.e. non-material, non-concrete and general, e.g. if the cue word length has a value of 7, then this includes all words which contain 7 characters, irrespective of their physical length. Likewise, the presence of crosses means that the word contains one or more of the characters 'f' and 't', irrespective of the number, location, or style of the physical marks (see section 2.2.8 and section 2.4). The cues are more relevant to identity than to pattern, e.g. the letter 'a' identifies every written instance of 'a', regardless of form (see section 2.4). The cues are used to search a lexicon and the results therefore reflect lexical, rather than physical, properties of the word, i.e. the kind of word patterns which are present in a lexicon and the way in which the cues partition a lexicon (see section 4.2). The pattern of cues does not correspond to the physical form of a word, instead they correspond to an abstract representation of a word, e.g. the words "abnormality", "affectingly" and "assertively" have the same abstract representation.

The word shape recognizer, like the pattern recognizer, places the same reliance upon all of the letters of a word, irrespective of their position. This is not the case with the word level method. The word level method uses the cues first letter and last letter because, for human readers, end letters are more important than middle letters. The word level method is not attempting to recognize the shape of a word. It uses general, abstract cues not physical features, and it draws upon contextual evidence, not physical evidence.

6.8 Experiment 13: Confirming the Word Level Method

6.8.1 Introduction

In Chapter 3 the hope was expressed that the machine system could be as competent as human readers if it could use word level cues. The previous experiments above had shown that it was possible to use the word level cues to generate a set of word candidates. An experiment was conducted to discover what improvement in performance the addition of the word level method and letter verification caused. This experiment will also show that the word level method is robust.

6.8.2 Method

The data set that was used in Experiment 1 was again used in this experiment, that is, 26 lower case words, with each letter of the alphabet contained in at least one of the words (see section 3.3). The procedures used for deriving word level cues and applying the word level method were exactly the same as used in experiments 8 to 11 (see section 6.2 to section 6.5). That is, no adjustments were made to accommodate the data set being tested.

Each word was presented to the recognition software and its response recorded. Word level cues were derived using the procedures set out in Chapter 5. The word level method was used to generate a new set of candidates using these cues. The confusion matrices and probabilities which had been created in Experiment 7 were used to generate these candidates (see section 6.2). The resulting word lists were merged with the original output from the pattern recognizer. Lastly, the letter verification recognizer was applied to the merged lists.

6.8.3 Results

Table 6-2 shows percent correct recognition for the human subjects, the pattern recognizer, and for the integrated system (the pattern recognizer merged with the word level method, plus letter verification). The column results are, respectively, correct recognition of the target letter, correct recognition of the target word and, lastly, correct recognition of the target word or, if the word has not been recognized, correct recognition of the target letter in the word (see section 3.2.2, and Table 3-1). Note that the machine results reflect the top ranked choice.

	letter	word	letter in word
human	75%	85.6%	93.9%
pattern recognizer	65.4%	61.5%	73.1%
integrated system	65.4%	69.2%	84.6%

Table 6-2: Percent correct recognition by human, pattern recognizer and the integrated system

The word level method improved machine performance (from 65.4% to 69.2%). However, the recognition system is still not as competent as human readers.

The word level method has been applied both to a test set and, in this experiment, to a second data set. Confusion matrices and probabilities derived from the training data have been applied to two different data samples. This confirms that the improvement caused by the word level method is a robust one. It has been shown that the word level method is effective.

6.8.4 Discussion

It is important to be clear that no claim is being made that the word level method is the same as the WSE. The word level method does not act in the same way as the WSE and the word level method is not meant to be a computer simulation of the WSE. For instance, no explanation for the WSE has been agreed upon. The cues which are used by the word level method are not necessarily those used by human readers. Whilst it may not be agreed that orthographic regularity, phonological regularity and morphological information play a part in the WSE, the word level method makes no use of such information at all.

However, there is compelling evidence that human readers use some of the cues used by the word level method, i.e. word frequency, word length, first letter and last letter (see section 2.2.1 and section 2.2.3). Justification for the use of the other cues has been presented in section 2.2.8 and section 2.4. It has already been demonstrated that the pattern recognizer shows a slight improvement at the word level as a result of lexical constraints (see section 3.2.4). The improvement in performance caused by the word level method is, therefore, above and beyond any lexical effect.

The improvement in performance caused by the word level method is not just quantitative, but also qualitative. The characteristics of the output of the integrated approach are different to those of the pattern recognizer. The distribution of the results of the integrated method is different to the distribution of the results of the pattern recognizer. Results show an increase at all ranks, but this increase is particularly marked for lower ranks.

The human subjects recognized words significantly better than the letters alone, and were most proficient with letters in words. The pattern recognizer performed worse on words than letters alone, but performed better on letters in words than letters alone. The difference between letters and letters in words was much smaller than that for human subjects, and the difference between letters and words was in the opposite direction. In contrast the integrated system displays the same characteristics as human readers. The integrated system recognized the target words significantly better than the letters alone, and its best performance was with letters in words. The difference between letters and letters in words was close to that for human subjects. The human subjects showed an improvement from 75% letter to 93.5% letter in word, which was a rise of approximately 20%. The integrated system showed an improvement from 65.4% letter to 84.6% letter in word, which was also a rise of approximately 20%. A context effect at the word level has been observed. Like the WSE, the word level method uses contextual cues to increase recognition performance at the word level. The integrated system recognizes letters in words more accurately than letters in isolation and this is, in essence, the WSE.

6.9 Conclusions

The word level method exploits the same information as the pattern recognizer but exploits the information in an entirely different way. The word level method consequently displays different characteristics to that of the pattern recognizer. The word level method lacks the fine-grained recognition abilities of the pattern recognizer. It is very poor at picking out the target word, but is likely to include it amongst the candidates. Further, because it is based on lexical rather than physical properties, it chooses different words than the pattern recognizer, so that when the two methods are taken together, a successful outcome is more likely. The way to create a machine system which is reliable and suitable for post-processing is to combine different, but complementary, recognition methods.

The word level method uses contextual cues to increase recognition performance at the word level. The word level method does produce an increase in machine performance at the word level which is similar to that of the Word Superiority Effect. It applies contextual cues to select from a lexicon. The word level method is primarily a top-down method which uses abstract word level cues. The word level method has been integrated with a bottom-up approach towards recognition. Combining the results of the word level method and the pattern recognizer provides a first stage for integrating high-level and low-level information.

An important point about the word level method is that it is an early attempt to develop an integrated system. The word level method can be considered a higher level process which, significantly, adds new candidate words to be considered for recognition. This is different from other higher level processes which have been explored [e.g. Evett, et. al., 1992], which merely select from word candidates proposed by a pattern recognizer. It is hoped that integration at this level will facilitate integration with higher level knowledge. It is clear that such integration is necessary for robust performance.

Chapter 7: Discussion

7.1 Discussion

Cursive script recognition is problematic because of the great variability between writers and in the writing of a single individual as well as the difficulties in segmenting characters. The human recognition of written text is not solely dependent upon pattern recognition. Rather a number of different sources of information, together with the reader's knowledge about written language, are used in conjunction with character recognition during the reading process [Monsell, 1991; McClelland, et.al., 1992].

Human readers recognize letters in words more easily than letters in isolation [Cattell, 1886; Baron & Thurston, 1973]. This is a context effect at the word level. One great advantage which human readers have over a machine system which relies purely on pattern recognition is their ability to take into account contextual information at the word level. This thesis has explored methods for deriving and integrating word level information into the recognition process. A number of different sources of information, such as word frequency and word length, have been chosen as being of use at the word level. This selection was made on the basis of research carried out within the field of cognitive psychology.

This thesis suggests some factors which can be used to improve the performance of a machine system. Preliminary work on the integration of information taken from the meta-word level has been described. There is an upper limit to the accuracy of any pattern recognition system. It is possible to predict the performance expected at the letter level and, by implication, the degree of improvement expected at the word and meta-word levels.

The WSE is a robust effect in printed text, Experiment 1 confirmed its presence in the reading of cursive script (see section 3.2). Human readers are nowhere near perfect on recognizing a set of letters. The machine system is worse than humans, although not far behind and better than one of the human subjects. It is apparent that the recognition system will have to exploit sources of word level information if it is to be as capable as humans. Experiment 1 also showed that the pattern recognition system developed at the Nottingham Trent University is performing at a level comparable, if at the low end of the range, to that of human readers in the recognition of letters in isolation. The exploitation of word level cues, as demonstrated in Chapter 6, improves the performance of the machine in the recognition of whole words, in particular when letter verification mechanisms are also added. However, Experiment 13 indicates that the machine system is not yet as accurate as human readers (see section 6.8).

The experiments presented in Chapters 4 and Chapter 6 demonstrate it is sometimes possible to use the set of word level cues to identify the target word even though the pattern recognition software has completely failed to give the target word as an alternative. In those cases where the pattern recognizer failed to give the target word as an alternative, 18% of the target words were top ranked by the word level method, 42% were in the top 10, and 64% were in the top 100. In those cases where the pattern recognizer did not place the target word top ranked (including where the target word was not given as an alternative), 27% of the target words were top ranked by the word level method, 58% were in the top 10, and 74% were in the top 100. The accuracy of the two methods differ. However, different kinds of accuracy have been identified. The pattern recognizer is maximized for a particular purpose: choosing the target word as the top ranked choice. It is relatively efficient at fulfilling this purpose. No surrounding contextual information is assumed. The pattern recognizer is thus discriminatory. The pattern recognizer can be considered to be highly discriminatory because it can often specifically identify the target word; it tends to place the target word at the top of the ranked list of word candidates. However, the pattern recognizer is not always successful. The pattern recognizer is fragile: it tends either to get the target word correct (to place it at the top of the alternatives), or to completely fail to identify the target word. There are relatively few cases in which the pattern recognizer identifies the target word but places it at a lower rank.

The word level method is, by contrast, less discriminatory than the pattern recognizer but more robust, e.g. one poorly written character can cause the recognizer to misidentify a word, but will not necessarily cause the word level method to fail. The word level method can also deal with poorly delineated characters, i.e. it bypasses some of the problems involved in segmentation. However, the word level method is often unable to precisely identify the target word within the set of word alternatives which it has generated. The word level method is therefore more robust, but less discriminatory, than the pattern recognizer. It has been demonstrated that it is possible to successfully merge the outputs of the pattern recognizer and the word level method (see Chapter 6). The consequence of integrating these two methods is a system which is both discriminatory and robust. Two recognition methods which display significantly different characteristics can be integrated to improve machine performance. Experiment 10 shows that letter verification provides the kind of additional support needed by the word level method to increase the probability that the target word appears towards the top of the list (see section 6.5).

It is not possible to get complete disambiguation from pattern recognition alone since a written word can be interpreted in a number of different ways. The combination of several sources of information, each of which is capable of extracting a different characteristic of cursive handwriting, is more likely to be successful. The way to produce a machine system which is both discriminatory (target top ranked) and robust (target found somewhere) is to combine different, but complementary recognition methods. It is only by integrating different sources of information that a stronger, more accurate, machine system can be developed. Two sources of information have been considered: character-segmentation information, and word level information. The methods used to extract these two sources of information are, respectively, a conventional pattern recognizer, and a method which uses word level cues.

A word level method has been developed. One source for word level cues is the candidate list generated by the pattern recognizer (see Chapter 4 and Chapter 5). The word level method uses, in this case, information which is already present in the candidates suggested by the pattern recognizer, but structures and re-organizes this information. It also applies it in a different fashion to that of the pattern recognizer. Word level cues are used to derive a new list of alternatives to add to the existing ones generated by the pattern recognizer (Chapter 4 and Chapter 6).

A number of cues which are useful at the word level have been set out. These cues are used in conjunction with, firstly, knowledge about probable confusions and, secondly, knowledge of the kind of word patterns which are present in a lexicon (see Chapter 6).

A way to derive values for these cues using the list of candidates has been developed. An alternate source for these values is direct pattern recognition (see Chapter 5). It is possible to combine these two sources to improve the accuracy of cue detection. The word level method tends to generate the intended word, but requires additional support (e.g. word frequency information and letter verification) in order to increase the probability that the intended word appears towards the top of the list.

Experiment 11 looked at the integration of word level information and syntactic information into the recognition system (see section 6.6). This experiment showed that integrating word level information and context effects into the recognition process is effective. For instance, increased recognition by the machine even at lower rank shows the benefits word level information can have to the machine since it provides the correct value to be selected using syntactic and semantic information. The method of using word cues to derive a list of possibilities is analogous to the human reader's ability to derive suitable candidates even from poor handwriting.

It is argued that the application of word level information in the recognition process will mean that the user is provided with better and more appropriate cues when misrecognition occurs. It is a characteristic of on-line machine systems that they tend to provide the user with few appropriate indications as to why the machine system failed [Wolf, 1990; Frankish, Morgan & Noyes, 1994]. A machine system which fails to recognize a word may produce a top ranked choice which has only a slight resemblance to the intended word. It has been suggested above, that in the case of the pattern recognition system developed in the Nottingham Trent University, this phenomenon can be the result of quite minor cases of character misidentification since lexical lookup will, in turn, cause whole-word substitution errors. Frankish, Morgan and Noyes observe that the

> ...user's expectations of handwriting systems are usually based on their intuitive knowledge of the human recognition process. Recognition failures are interpreted in terms of static, rather than dynamic attributes of handwritten text. When diagnosing the causes of misrecognition, users inspect the 'ink' trace, and compare this with canonical letter forms... Firstly, confidence in the recognition process can be undermind by substitution errors that bear little visual resemblance to ink traces. Secondly, attempts at re-entry of misrecognised characters may fail because users' incorrect model of the recognition process causes them to modify their writing patterns in inappropriate ways. [Frankish, Morgan & Noyes, 1994]

Machine systems can provide the user with substantive mis-cues. These mis-cues are the result of differences between human and machine recognition processes. A machine system should provide the user with some understanding as to the causes of misrecognition and a user should have the opportunity to modify his or her writing and so re-enter a misrecognized word. The cues used by the word level approach have been derived from the psychological literature and the approach to recognition resembles the human recognition framework. Since it is the case that human readers strongly rely on visual cues such as first letter, last letter and word length, then a response by a machine system which is particularly sensitive to such cues will be more appreciated by the user and will provide the user with more appropriate information as to how he or she should modify his or writing.

It was argued that the pattern recognizer needs to be more sensitive to the requirements of contextual analysis. It is necessary for the pattern recognizer to provide a better indication of its certainty of the candidates which it has chosen. The recognizer also needs to be more sensitive to the legibility of the input. It can be shown that writing speed effects legibility [Suen, 1983]. It is possible to determine writing speed from on-line handwriting. It may be possible to use this information to estimate an expected level of legibility.

It is necessary to identify and explore the different machine applications for a script recognition system. This is important since some of the sources of information which can be used, and the way in which these sources can be integrated, is dependent upon the application. One possible application is a note taking system for doctors. This kind of application is characterised by the recognition of single words in a highly constrained domain environment. The requirements of such an application will obviously be very different to a larger text recognition system. The extension of script recognition to full text comprehension needs to be examined. This will be useful since it will highlight the possible limits which may exist to the machine recognition of cursive script but in particular it will require the creation of a complete system in which all of the various recognition methods which have been developed will necessary.

7.2 Further Work

A machine system must take both context as well as pattern recognition into account if it is to be effective. A method for deriving and exploiting word level cues has been presented. The addition of syntactic and semantic information into the recognition system has also been shown to be cogent [Keenan & Evett, 1994; Rose & Evett, 1993]. It is proposed that the various methods for utilizing contextual information be built into a unified machine system. Some ways in which different sources of information can be integrated will now be discussed. The possibilities offered by integration for the development of combined bottom-up and top-down mechanisms within a unified system will also be examined.

Integration has an important place in the machine recognition of unconstrained handwriting [Evett, et. al., 1994b]. The integration of different sources of information will allow a machine system to compare different sources of information in order to verify its choice, to attend selectively to the text, to look selectively for key words and phrases, and to anticipate what will come next. Integration will make the pattern recognition stage much less critical and since pattern recognition cannot hope to be wholly accurate, this can only be of benefit.

Integration will serve three ends. Firstly, integration will improve machine performance by making the machine system more accurate. Integration will improve accuracy since it is possible to use contextual information to select from the list of alternatives generated by the pattern recognizer. The combination of several different sources of information will therefore improve machine performance.

Secondly, integration will make the machine system more robust. The word level adds new candidates to the list of alternatives provided by the pattern recognizer. The amalgamation of the word level method with the pattern recognizer makes for a more robust system since the word level method shows a greater tendency to provide the target word as a candidate, even though it is not always able to place the target word towards the top of the list of alternatives. To put this the other way around, the introduction of the word level method makes it less likely a target word is unrecognized.

Thirdly, the combination of different sources of information offers the opportunity to combine both bottom-up and top-down processes within the system. In the current system the pattern recognizer (the starting point in the system) guides the higher level processes by providing the set of word candidates from which the higher level processes are forced to select. The introduction of the word level method may add to this set of word candidates but information still flows only one way: from the bottom to the top. It is possible for higher level processes to influence or guide the operation of lower level processes. For example, it is possible for a higher level process to initiate a more detailed analysis of a word by a lower-level process, or for a lower level process to verify selected candidates [e.g. Hull & Srihari, 1986; Lecolinet & Crettez, 1991; Simon, 1992; Madhvanath & Govindaraju, 1997].

The current system is only integrated in a simple fashion. Current work is aimed, firstly, at making the integration more successful, secondly at adding a further layer into the system (the word level method) and, lastly, developing communication between the different layers of the system.

Each source of information drawn upon during the reading process acts as an aid to recognition. In the existing system the various higher level processes do not add to the original set of letter candidates produced by the pattern recognizer. Rather, sources of higher level information (lexical, syntactic and semantic) are only used as a series of filters which act to reduce the set of letter and word candidates [Wells, et. al., 1991]. For example, candidates are ordered on the basis of syntactic considerations. Meta-word information is therefore used to constrain the candidate set, the net effect of this being a filtration process. In other words, the context in which a letter is presented only influences the accuracy of the post-recognition process, not the process of letter recognition itself. The flow of information is only from bottom to top. Filtration is simple and fast, but it does not make full use of higher level information. The integration of the word level method into the machine system means that for the first time, except for the initial candidate set generated by the pattern recognizer, a part of the system is being allowed to expand the candidate list.

It is apparent that for human readers some form of interaction between letter recognition, and lexical, word, syntactic, semantic and pragmatic information is vital to recognition. There is a strong relation between knowledge and perception. Knowledge about words, sentences and meaning influences the process of perception. Script recognition is a prime example of a dynamic system using a feedback mechanism

Different approaches towards the recognition of cursive handwriting have been identified. In the first approach the recognition process is driven by the pattern recognizer. This is a bottom-up approach. The second approach uses the lexicon and applies contextual cues to select from this list of words (e.g. the word level method). This is a top-down approach. The reason why the latter is desirable is that some words are not well written and, in such cases, it will find the word whilst the first method will not. Integration also offers the opportunity to combine bottom-up and top-down approaches within the system, i.e. the use of letter verification procedures.

However, it is also hoped that integration may have other benefits as well. Given sufficient knowledge about contextual cues such as a word's grammatical class, semantic relevance and certain word level cues it becomes possible to identify the word without having to engage in any more detailed analysis of its individual constituents [Rayner, et. al., 1982]. Integration may therefore reduce or even remove the need for extensive pattern recognition

The integration of different sources of information is not a trivial problem. The reason for this is that the characteristics of the textual knowledge offered by these various sources of information are dissimilar and the kind of data which can be derived from the different sources is qualitatively different. It must also be possible to evaluate the strength of the different sources of information. This is particularly important in those cases when the interpretation suggested by the different sources of information conflict. Whatever mechanism is adopted for integration must therefore be able to compare and contrast qualitatively different sources of information [Xu, et. al, 1992].

One way a method of handwriting recognition can be judged is on the basis of the target word being top ranked. Such a judgement calls for one forced choice. However, when a method can be used in combination with other methods, then it may be more important that the intended word is found, even if it is not given as the top ranked choice. The criteria used to evaluate methods of handwriting recognition will depend, in part, on the context in which it is to be used.

The purpose to which a recognition system is put will, to a large extent, determine the sources of information which can be used and the way in which these sources are integrated. There are many applications for script recognition, some of which require different machine capabilities and therefore different system configurations. For instance, a notepad application might require only single words to be recognized. Since meta-word contextual information is not available then it is not possible to use syntactic or semantic analysis. However, it may be possible to use pragmatic information with such applications, e.g. if the purpose of the application is to allow a user to complete a form then knowledge about the various categories within the form could be used to tailor the lexicon to each particular category. If text is to be recognized then syntactic and semantic information can be used. However, semantic information can sometimes be exploited even when textual information is not available. For example, knowledge of the domain can be used to aid final selection even though the application calls for single words to be recognized in isolation.

The point where integration is made during the recognition of text is also of interest. There is debate about whether human readers use all possible sources of information immediately (i.e. word by word) or whether these sources are only brought together at the end of a sentence or clause. If a machine system is to operate in real time then only the first option is available. One would expect a user to notice and correct a misidentified word straight away rather than be willing to continue writing in the hope that the machine will correct the word itself given more information. However, there are applications, such proof reading and Optical Character Recognition, where it is possible to examine the overall text. This will be of benefit to the recognition system since more information will be available during the recognition of any one word.

One important issue in integration is the interaction of effects. One possible side-effect of combining two methods is that the same effect may be present in both methods. This is particularly important once interaction is introduced. There can be an overlap between the set of cues which are used by one method and the set of cues which are used by another, e.g. the presence of ascenders. This kind of repetition is not necessarily that obvious. For instance, there is a co-relation between syntactic class and word frequency. The lexicon which has been used in the experiments described in this thesis is itself based on word frequency. It may be the case that an unintentional word frequency bias exists in several of the methods which have been developed. The effect of such kinds of unintentional repetition will only become apparent when interaction between different methods occurs in the integrated system. It is possible to see that destructive forms of feedback could occur in the system. It is possible for unforeseen side-effects to arise even though a particular method, examined in isolation, appears both efficient and accurate.

Currently, only the word level method and the pattern recognizer create new word alternatives. One possible approach is to also allow meta-word level processes to add to the word list. For example, the system is attempting to read the sentence, "He saw two dogs". However, the 's' of "dogs" has been poorly written and the letter has not been suggested as a candidate by the pattern recognizer. In this case syntactic knowledge can be used by the system to indicate that the singular "dog" should actually be plural, and therefore to insert the correct letter into the sentence. Note that the 's' has been neither suggested nor verified by the pattern recognizer but rather that a letter has been inserted by a higher level process.

An alternative approach is to allow higher level processes to influence the operation of lower level processes. In this second case the example sentence will be analysed as it is read in, and before the final character is reached the system will be indicating that the word is likely to be plural and hence the pattern recognizer's ability to identify the final letter as an 's' can be improved. Probabilistic semantic and syntactic analysers have a predictive capability. Markovian approaches to implementing syntactic information are such that the probability of syntactic class for any word position can be estimated on the basis of its word context [Keenan & Evett, 1994]. Evaluation of the use of probabilistic syntactic information has been carried out. Probabilistic semantic information can also be applied [Rose & Evett, 1993]. The first approach outlined above (adding to the word list) may have its place in a script recognition system (e.g. to cope with spelling mistakes), but it is suggested that this second method will be more accurate and that it offers the possibility of developing a fully interactive system.

The application of syntactic and semantic constraints means that probable candidates from a word list are selected. Currently, this selection process is only used to guide the final output of the system. However, once selection has been carried out, it is possible instead to initiate a more detailed analysis of the selected word or words. This is an obvious option when higher-level processes have been unable to make a choice between candidates, but the introduction of an automatic verification procedure might also be desirable. This kind of analysis can be done by the pattern recognizer or at the word level. At the word level a more detailed analysis of the word could involve a refinement of the cues already used, e.g. to count the exact number of ascenders in the word rather than just whether ascenders are present or absent.

Experiment 10 demonstrated one mechanism for combining top-down and bottom-up approaches (see section 6.5). In this case, the recognizer was asked how many characters from each of the candidates it could identify in the ink pattern. This is a request for confirmation by the pattern recognizer of the presence of particular characters. For example, a significant number of those cases in which the word level method had difficulty identifying a word were the result of alternative spellings of the same word, e.g. recognize/recognise. The method could be further refined so as to direct attention in these cases to the specific letter (or occasionally letters) which cause these alternative spellings.

It is possible to use a pattern recognizer to verify selected candidates in two other ways. Firstly, the recognizer could be told to assume that a particular character does in fact exist. The recognizer would then attempt to see whether it could match the other characters to the predicted word.

Secondly, it is possible to extract further letters or letter sequences from the word. Such an approach focuses upon the sequence in which particular the characters occur. For example, two letters which are frequently confused are ‘y’ and ‘g’. However, when the letter ‘g’ is placed at the end of a word it typically occurs as part of the letter sequence "ing", whereas ‘y’ is most commonly preceded by the letter ‘l’ but only very infrequently as "iny". See Table 7-1. Similar differentiating character sequences can be shown for the other commonly confused letters. Directing a search through the word in this manner has obvious benefits.

letter sequence	letter sequence in lexicon
'g'	11.37%
'ng'	11.29%
'ing'	11.21%
'lg'	0.0%
'y'	10.59%
'ny'	0.13%
'iny'	0.02%
'ly'	6.49%

Table 7-1: Examples of the proportion of words in a 15k lexicon containing particular end letter sequences

The pattern recognizer, word level method and meta-word level methods differ with respect to the kind of input used, the type of information generated, and the degree and kind of discrimination which the method has.

The pattern recognizer and the word level method both produce an ordered list. Each word in the list has a confidence score associated with it. One target word at a time is processed. Input to the word level method from the pattern recognizer is in the same format as output. Semantic and syntactic analysis use an analytical method which is not dependent upon the ordering of the word list. Multiple word lists are used to construct a word lattice. In the current system, the results of the analysis carried out by each of the methods is then combined with the scores given to each of the candidate words by the pattern recognizer. The lattice is analysed using the probability of a tag appearing in the given context for the syntactic analyser, and how well words combine with neighbouring words in the input for the semantic analyser. The current candidate list is then ordered based on this combined analysis, i.e. output is an ordered list.

Although unordered lists are used to construct a word lattice, one problem associated with the use of a word lattice is combinatorial explosion. The number of routes through the lattice (assuming a uniform list size) is equal to the number of rows to the power of the number of columns, e.g. a word lattice constructed out of 3 word lists each of which contains 4 word alternatives will generate 43 = 64 routes through the lattice. In practice, therefore, a threshold based on the ordering of the input list is used to limit the number of words used, currently this threshold is set at 10 words. An alternative would be to optimise routes through the lattice using probability [e.g. Viterbi, 1967]. This would overcome the necessity of such a threshold. Meta-word contextual methods are less efficient if the list of word alternatives is too long. For both syntactic and semantic analysis, the greater the lattice, the lower the average accuracy rate, i.e. there is a decrease in discrimination because of the increased likelihood of spurious connections.

The pattern recognizer currently lacks the capability to make a strong distinction between clear and unclear handwriting. It may be possible to give it this capability. However, physical cues should hold sway over contextual cues. It appears to be the case that for human readers contextual information is subordinate to perceptual cues [Bouma, 1971]. It is only in the case of badly formed or ambiguous script that contextual information should override pattern recognition. The less clear a piece of writing the more reliance has to be placed upon contextual cues. It is when confidence at the pattern recognition level is either low, or when the pattern recognizer has given two or more candidate words the same confidence score, that contextual cues are of particular benefit.

The word level method is less discriminatory than the pattern recognizer, but it also less fragile. The word level can be used to gain an improvement in the number of target words top ranked but its main effect is to increase the number of target words found regardless of rank. The pattern recognizer and the word level method can be successfully combined so as to produce a system which is both discriminatory and robust.

The kind of information derived from syntactic and semantic analysis is less discriminatory than that derived from the pattern recognizer, e.g. syntactic analysis indicates that a particular class of words are appropriate, rather than a particular word or small group of words. For instance, many English words are syntactically ambiguous. Broadly speaking the grammatical classes which are used most frequently are nouns, then verbs. Approximately 45% of words in the Oxford corpus are nouns. It is possible to partition a word list into probable, less probable and improbable groups of words using syntactic analysis, but it is rarely the case that such analysis can, on its own, select just one word as being appropriate. The same is true of semantic analysis.

Pattern recognition is not related to word frequency in any way. Its only consideration is the input pattern. However, the word level method is supported by word frequency information therefore recognition is biased towards higher frequency words. There is also a partial relationship between syntactic class and word frequency, e.g. function words tend to have a very high frequency of occurrence. Words which occur less frequently tend to belong to grammatical classes which have a high frequency of usage, e.g. nouns and verbs. Higher frequency words are not semantically restrictive. For instance, around 50 words account for about 30% of the Oxford Corpus. All of these words are function words. Higher frequency words manifest few 'selectional restrictions'. For example, the word "on" is used to describe physical relationships but also used to describe metaphorical relationships. It is not that the word has multiple meanings, but that it can be successfully used in relation with a large number of other words. Probabilistic methods of semantic analysis are not suited to the task of disambiguating function words, so other methods have to be employed (e.g. syntactic analysis and pattern recognition methods).

Any reduction of the lexicon should mean an increase in recognition performance. It is possible to use syntactic analysis for lexical reduction. It is suggested that an initial parse for function words be carried out. Function words are a small, closed class of words which perform a specialised job in the English language. Function words clarify the relationships between content words. They join content together. Function words have a very high frequently of occurrence in written text. Function words are well established in a language. It is exceptionally rare for a new function word to appear in a language. Function words are few in number. The exact number of function words, however, is undecided. Different writers specify different sizes to the group of function words, depending on the definition used. However, a consistent set of words appear in any list. These include such grammatical classes as articles ("the", "a", "an", "this"), prepositions ("in", "on", "until") and co-ordinate conjunctions ("and", "or", "but"). A small, probably only 200 word, list of function words is envisaged. Function words belong to very specific, sometimes unique, grammatical classes. Syntactic analysis can therefore provide particular support to the recognition of function words.

It is possible to match a specific set of words to the particular methods most suited to their recognition. For example, function words tend to have a short word length and a very high word frequency. This means that it should be possible to exploit the particular characteristics of the word level method in order to recognize this specialised group of words, e.g. it has demonstrated a high accuracy in the prediction of short word lengths. The role which function words play in English means that there is little or no chance of using semantic information as an aid to their recognition. Since function words have a high frequency of occurrence in written language then successful identification will improve the average recognition rate for text. This should also help to bolster user acceptability.

There are other sources of information, apart from those already indicated, which might be of use to the machine recognition of script. For example, Ranklin suggested that word length sequence could be a reading cue [Ranklin, 1977]. Noun and verb phrases often take the form of ascending word length. This kind of word length structure probably has syntactic origins. Since the word level method provides good estimations of word length, it is possible to explore the usefulness of this structural cue. An area which has not yet been explored is the recognition of punctuation marks and paragraph breaks. Clauses tend not to contain a great many words. Even in writing, clause structure is often determined by breathing patterns, i.e. the number of words which can be comfortably spoken using one breath. Another factor which may limit the size of clauses and sentences is memory, e.g. it is more difficult to remember the topic of a sentence which is overly long. Kucera and Francis calculated that the mean length of a sentence was 19.2 words [Kucera & Francis, 1967]. Very long sentences are unusual since they are difficult to understand. For instance, greater sentence length adversely affects reading time. All of these kinds of information can be labelled pragmatic. It should be possible to use this information, e.g. in the prediction of punctuation marks.

If the system is not confident in its ultimate choice then word frequency can also be used to make a final selection. Word frequency can also be used in those cases where all other methods of disambiguation have failed, i.e. when the system is undecided. High frequency words will obviously tend to occur more often in the input data. Applying word frequency to the word list will have the effect of probabilistically increasing the number of target words placed as the first choice in the word list. This is purely a statistical gain: the average number of times a target word is top ranked. However, human writers will probably find a failure to recognize low frequency words less irritating than a failure to recognize high frequency words and the system will more correct more often.

A further approach which is under consideration to build relevant, non-physical factors, into the recognition process; such as letter frequency and orthographic and phonological regularity. This is not simple task however; the system must still be able to recognize irregular words when they occur and a method such as Markov models would have to be employed to integrate these sources of information [e.g. Shinghal & Toussaint, 1979; Farag, 1979; Ford & Higgins, 1990].

A method of communication between the different layers of the system has been set out during the course of Chapter 6. This method uses lists of word alternatives and confidence scores for each of these alternatives. These word lists are merged together and the confidence scores of the candidates are averaged. Three different sources of information have been used in this thesis: a conventional character-based pattern recognizer, a word level method, and a letter verification recognizer.

The way in which different sources of information fit together in the human reading process is not yet fully understood. It cannot be precisely determined how close any process within a machine recognition system corresponds to its human equivalent. It is therefore necessary to empirically test different configurations of the machine system in order to determine their efficiency. Estimates of the accuracy of each method can be made. Since each method attempts to use a particular source of information, it is also possible to indicate how important these sources of information are to human readers and the reliance which readers place upon them in different circumstances. However, the weight given to each method when it is placed within the context of an integrated system will, again, have to be decided empirically. Specifically, empirically based decisions will be made about the degree of initiative allowed to each source of information within the context of the whole system, the level of influence which each source is allowed to exert, and the form, strength and purpose of the information passed onto neighbouring processes.

A working, integrated, machine recognition system has been implemented. This system combines word level cues with a conventional character-based pattern recognizer. The word level method has proved successful. The word level method has created a context effect at the word level. The word level method combined with the pattern recognizer performs better, in all respects, than the pattern recognizer on its own.

Nine sources of word level information have been used in this work: seven word level cues and two higher level contextual sources (lexical and word frequency information). These cues were selected because they are effective sources of information at the word level. However, selection was also guided by what could reasonably be obtained from the pattern recognizer. The development of new recognizers will allow the exploitation of alternative cues. The cues presence or absence of ascenders, descenders, dots and crosses were chosen because they are a set of generalized word level cues which preserve the sort of information retained across letter confusions. It is apparent that there are other cues which could have been chosen instead of these four cues, for example, holes and curves. Whilst the chosen cues have proved successful, further work to test the usefulness of other cues is obviously called for. It is not foreseen that other cues will be more successful than the chosen cues, rather these alternative cues will serve as additional sources of information and hence improve the overall effectiveness of the word level method.

A number of sources of information other than those used by the word level method were described in Chapter 2, for example, phonological regularity and morphological information (see section 2.2.7). It is not agreed that these sources of information do play a part in the Word Superiority Effect. It was for this reason that these sources were not used in the current work. However, there are indications that these sources of information influence human reading. Further work is therefore needed in order to examine these other sources of information and to explore ways in which they can be included within the word level method. Orthographic regularity is of particular importance because the machine recognition system will not be limited to a given lexicon as currently. The implementation of such an approach will obviously be a major task, but it will result in a more flexible and general recognition system. However, such an approach may be less accurate than the use of a fixed dictionary [Ford & Higgins, 1990].

Further work on the lexicon is called for. The use of different lexicons should be examined, e.g. developing a core vocabulary for the system. A core vocabulary is made up of high frequency words. These are the words which, together with function words, make up the bulk of written language. In the current system the lexicon was created using the 15,000 most frequent words. The use of a different lexicon may affect the way in which the word level method uses word frequency.

Further work on improving of the word level method is indicated. The current work has shown that some methods and values are better than other methods and values. However, it is apparent that alternative parameters should be explored. For example, one consequence which follows on from the introduction of new methods is that the set of parameters used, and the combination of methods employed, will have to be re-examined. The task of optimizing the many parameters used to derive values for the cues, and then to apply these values, is complex. A great many methods, thresholds and ranges are involved in deriving word level cues and in the word level method itself. A way of automatically optimizing the parameters needs to be developed. Methods to carry out multi-parameter optimization therefore need to be examined, e.g. constraint satisfaction. The automation of multi-parameter optimization will, firstly, serve to improve the effectiveness of the word level method and, secondly, assist in the process of adapting the word level method to other recognizers. For example, the parameters used in the current work inevitably reflect the biases of the pattern recognizer being used.

The use of a data sample larger than the one used in the current work is called for. A larger data sample is necessary, firstly, because the word level method currently treats the cues first letter, last letter and word length as statistically independent. This means that the word level method is not providing as accurate results as is desirable. The reason for treating these cues as statistically independent is that the size of the data sample used to derive the probabilities employed by the word level method is not large enough. Secondly, the sample size was not large enough to generate reliable probabilities for some letter confusions, e.g. the letter 'x'. The word level method has been shown to be robust. It has been successfully applied to a test set and, in Experiment 13, to a second data set (see Chapter 6). However, a larger data sample will serve to improve the generality and flexibility of the current system.

Further work will seek to improve the way that the cues first letter and last letter are derived. The cues first letter and last letter are derived more accurately using the candidate list than using direct cue extraction. However, direct cue extraction is the only cue derivation method which can be used in those cases where the pattern recognizer has experienced a catastrophic failure. Direct cue extraction uses the letter graph generated during pattern recognition to calculate the most likely candidate for the first and last letter. One problem with the recognizer is that certain letters occur in the letter graph more frequently than warranted and they are ascribed too great a confidence score when they do occur. This is because such letters are 'simple letters': they are either easier to match with the patterns in the character database than more complex letters, or compose part of more complex letters, e.g. 'c' and 'l' are simple letters, whilst 'd' and 'y' are complex letters. Simple letters are more likely to produce a response from the recognizer than complex letters. Further work will seek to improve the performance of the recognizer on the cues first letter and last letter. For example, the confidence score of simple letters could be reduced, thus decreasing their influence on the calculations.

Increasing the performance of direct cue extraction for the derivation of the cues first letter and last letter will mean that it becomes possible to develop the word level method as the sole, or primary, recognition method. The reasons for this are, firstly, that the cues used by the word level method must be derived using direct cue extraction if the word level method is to be independent of the pattern recognizer. Secondly, direct cue extraction produced significantly worse results for the cues first letter and last letter than all of the other cues. These two cues, therefore, are the most important stumbling block in the way of using the word level method as the primary recognition method.

The word level method was not designed to be a stand-alone recognition method. Throughout the current work it has been assumed that the use of word level cues is not an alternative to the pattern recognizer. This assumption may be false. The word level method was designed to complement the pattern recognizer. At all stages in its development the word level method was evaluated on the basis of how well it would work in combination with the pattern recognizer. For example, methods were selected, and parameters were adjusted to support the pattern recognizer. The word level method can be used as a substitute recognition method on those occasions when the conventional pattern recognizer has failed. The results obtained from using the word level method as a substitute recognition method suggest that the word level method has the potential to be developed as the primary recognition method.

Future work will examine the effectiveness of the word level method as a recognition method in its own right. The model presented in the current work combines a conventional, character-based pattern recognizer with the word level method. The discriminatory powers of the conventional recognizer have been contrasted to the non-discriminatory characteristics of the word level method. The word level method requires a way to distinguish between candidates delineated by the same seven word level cues. Letter verification provides such a mechanism. In the current work the bottom-up approach of the conventional pattern recognizer has been combined with the top-down approach of the word level method. The word level method could be used as a first-stage recognition method which subsequently relies on letter verification to make a fine choice from the word candidates selected by the word level method. Such a system would be top-down driven, since processing would begin with word level cues and subsequently use letter verification as the main bottom-up approach.

Future work will therefore also examine ways in which the letter verification procedure can be improved. The task of improving letter verification procedures has an obvious connection with the task of increasing the performance of direct cue extraction for the derivation of the cues first letter and last letter since both tasks currently use information from the letter graph. However, it is possible to derive the cues first letter and last letter directly. Future work will therefore explore alternative ways of obtaining these cues, e.g. the development of new pattern recognition methods.

In the current letter verification procedure a word candidate is increased in rank on the basis of the proportion of the characters which the letter verification recognizer recognized within the handwriting sample. The existing method has a bias towards shorter word candidates. It is possible to remove this bias by taking account of the proportion of, and position of, characters recognized by the letter verification recognizer in the sample but not found in the word candidate.

Alternative mechanisms for combining knowledge sources could also be developed. For example, a useful approach would be to build separate modules to represent each source of information [Nadal & Suen, 1993]. Such an approach models the interrelation of information in a complex environment. Modules may work co-operatively or as adversaries, i.e. to produce a process of activation and inhibition using several sources of information in order to achieve a final choice [Bozinovic & Srihari, 1989; Fein & Hones, 1992].

Other kinds of contextual information which can be of use to the machine recognition of cursive handwriting have been identified. These include pragmatic information (e.g. sentence and clause length) and application information (e.g. domain). Further work is needed concerning the derivation of this information, and ways in which this information can be applied. The integration of new with existing sources of information and recognition processes will not be a trivial task. It is also a task which is strongly influenced by the application to which the script recognition system is to be used.

It has proved possible to merge information taken from the list of candidates generated by the pattern recognizer with information derived from direct cue extraction. A way to merge these two sources of information has been presented. However, the existence of two different sources of information means that it is possible to use one source to indicate the reliability of the information derived from the other. Further work, therefore, will use the known characteristics of the different sources of information to modify the way in which the information is used by the word level method.

The combination of the word level method and the pattern recognizer was more effective than the word shape recognizer, both on its own, and in combination with the pattern recognizer. A combination of the word level method and the pattern recognizer plus letter verification has proved even more successful.

Further work will examine whether a combination of all three methods will lead to further improvement.

It is possible to combine the word level method and the word shape recognizer in order to improve machine performance. Whilst a comparison between the word level method and a word shape recognizer has been presented previously, no attempt to integrate the word shape recognizer with the word level method was been made. Further work will explore two ways in which the two approaches can be combined. Firstly, the word level method can be combined with the word shape recognizer using the same techniques that have been used to integrate the word shape recognizer with the pattern recognizer. Secondly, the word level method can be used as a pre-processor for the word shape recognizer. The reason for using this particular method of integrating the word level method with the word shape recognizer is that a smaller lexicon increases the accuracy of the word shape recognizer. The word shape recognizer can use the list of words generated by the word level method as its lexicon.

Once the most effective way to combine the word level method with the word shape recognizer has been discovered then it will be possible to explore ways to combine the word level method, the word shape recognizer and the pattern recognizer. A combination of all three methods should lead to further improvement.

Work on the integration of the word level method with a conventional pattern recognizer has been presented in this thesis. This work demonstrates that two different recognition methods can be successfully combined together into an integrated machine system. The consequence of integration is an improvement in machine performance. The accuracy and robustness of the machine system have been improved. A number of different sources of information have been drawn upon during the course of this work: output from a pattern recognizer, from direct cue extraction, lexical information and word frequency information have all been successfully combined. The combination of these sources of information serves to demonstrate how different methods can be successfully integrated and shows how a unified machine system can be developed.

Further work should examine the integration of meta-word information into the recognition system. Experiment 11 has shown that syntactic information will be of benefit to the machine system (see section 6.6). The amalgamation of syntactic information with the pattern recognizer and the word level method will draw upon work already completed [Evett, et. al., 1992]. Previous work has also shown that the addition of semantic information will be of benefit [Rose & Evett, 1992]. The various recognition methods display different characteristics. It is possible to exploit the characteristics of the different methods which are used in order to develop a robust, efficient and unified machine recognition system.

Syntactic and semantic constraints can be used to select from a list of candidate words. It is possible to use this selection procedure to determine the final choice of the machine system. However, it is also possible to use contextual information in an interactive fashion. Experiment 10 has shown the combination of top-down and bottom-up approaches within a unified system can be successful (see section 6.5). Syntactic and semantic analysis can also be used to initiate a more detailed analysis of a word. Future work can develop additional integrated processes, e.g. using meta-word contextual information to guide the conventional pattern recognizer or using pattern recognition to verify the chosen candidate.

The word level method has been solely applied to on-line cursive handwriting recognition. However, the approach to word recognition which has been presented in the current work is also applicable to off-line handwriting recognition and Optical Character Recognition. The specific word level cues which have been used in the current work are not necessarily applicable to these other problems. Similarly, the emphasis placed on the different cues will not necessarily be the same. There are different cues which are more readily identifiable in typeface than in cursive handwriting, and there are different cues which are more useful for the recognition of typeface than in the recognition of cursive handwriting. It has already been observed that there are other cues which could have been chosen in place of, or supplementary to, the cues used in the current work, e.g. holes. The approach taken in the current work towards word recognition will be relevant to any machine system which is attempting to recognize words. The success of the word level method on cursive handwriting strongly suggests that it is useful to pursue its application to other recognition problems.

The current work has also provided an insight into some of the characteristics of the pattern recognizer being used. For example, it has been shown that the recognizer is more accurate at determining some cues (e.g. the presence or absence of ascenders) than it is at determining other cues (e.g. the presence or absence of dots). The current work therefore has highlighted areas where further work would be effective and in which significant improvement to the pattern recognizer can be made.

References

L.R. Bahl & F. Jelinek, (1975), 'Decoding for channels with insertions, deletions & substitutions with applications to speech recognition', *IEEE Trans. Inform. Theory*, vol. IT-21, 404-410

D.A. Balota & K. Rayner, (1991), 'Word recognition processes in foveal and parafoveal vision: the range of influence of lexical variables', In D. Besner & G.W. Humphreys (eds.), *Basic Processes in Reading: Visual Word Recognition*, 198-232.

Paul Barber, (1988), *Applied Cognitive Psychology: An information- processing framework*, London and New York: Routledge.

J. Baron, & I. Thurston, (1973), 'An analysis of the word superiority effect', *Cognitive Psychology*, vol. 4, 207-228.

C.A. Becker, (1976), 'Allocation of attention during visual word recognition', *Journal of Experimental Psychology: Human Perception and Performance*, 2, 556-566.

G.J. Bellaby & L.J. Evett, (1994a), 'The Integration of Knowledge Sources for Word Recognition', in L.J. Evett and T.G. Rose (Eds.), *Computational linguistics for speech and handwriting recognition, One day workshop in the AISB 1994 Workshop series*, Leeds, April, 1994.

G.J. Bellaby & L.J. Evett, (1994b), 'The Integration of Knowledge Sources for Script Recognition', *Proceedings of IWFHR-IV, 4th International Workshop* *on Frontiers in Handwriting Recognition*, organising Committee Chair, Jhing-Fa Wang, National Cheng Kung University, ROC, pp 305-312, Dec 7-9 1994, Taiwan.

G.J. Bellaby, L.J. Evett & R.K. Powalka, (1996a), 'Multiple knowledge sources for word recognition', *Proceedings of IEE Workshop on Handwriting Analysis and Recognition- A European Perspective*, May 1996, organised by Professional Group E4 (Image Processing and Vision) pp 9/1-9/3.

G.J. Bellaby, L.J. Evett & R.K. Powalka, (1996b), 'Coping with ambiguity and error in script recognition', *Proceedings of IWFHR-V, 5th International Workshop on Frontiers in Handwriting Recognition*, University of Essex, England.

D. Besner & J.C. Johnston, (1989), 'Reading and the mental lexicon: on the uptake of visual information', In W. Marslen-Wilson, *Lexical Representation and Process*, Cambridge, Mass.: Bradford Book/MIT Press, 291-316.

G. Boccignone, A. Chianese, L.P. Cordella, A. Marcelli, (1993), 'Recovering dynamic information from static handwriting', *Pattern Recognition*, vol. 26, no. 3, 409-418.

U. Boes et. al., (1989), The Paper Interface: ESPRIT Project 295, Final Technical Report.

H. Bouma, (1971), 'Visual recognition of isolated lower-case letters', *Vision Research*, vol. 11, 459-474.

R.M. Bozinovic & S.N. Srihari, (1989), 'Off-line Cursive Script Word Recognition', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-11, vol 11, no. 1, 68-83.

D.E. Broadbent, (1967) 'Word frequency effect and response bias', *Psychological Review*, 43, 1-15.

M.K. Brown & S. Granapathy, (1980), 'Cursive script recognition', *Proceedings of the International Conference on Cybernetics and Society*, 47-51.

M.K. Brown & S. Granapathy, (1983), 'Preprocessing techniques for cursive script word recognition', *Pattern Recognition*, 16(5), 447-458.

D.J. Burr, (1983), 'Designing a handwriting reader', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-5, 554-559.

J.M. Cattell, (1886), 'The time taken up by cerebral operations', *Mind*, 11, 220-242.

D.W.J. Corcoran & R.O. Rouse, (1970), 'An aspect of perceptual organisation involved in reading types and handwritten words', *Quarterly Journal of Experimental Psychology*, 22, 526-530.

CEDAR database, available from CEDAR, SUNY Buffalo, USA.

K.W. Church, (1988), 'A stochastic parts program and noun phrase parser for unrestricted text', *Proceedings of the 2nd Conference on Applied Natural Language Processing*, 136-143.

R.G. Crowder, (1982), *The psychology of reading*, New York and Oxford: Oxford University Press.

R. Dodge, (1906), 'Recent studies in the correlation of eye movement and visual perception', *Psychological Review*, vol. 13, 85-92.

L.D. Earnest, (1962), 'Machine recognition of cursive writing', *Information Processing (Proceedings of the IFIP Congress)*, 462-466.

M.A. Eldridge, I. Nimmo-Smith, A.M. Wing & R.N. Totty, (1984), 'The variability of selected features in cursive handwriting: categorical measures', *Journal of the Forensic Science Society*, 24, 179-219.

S.F. Ehrlich & K. Rayner, (1981), 'Contextual effects on word perception and eye movements during reading', *Journal of Verbal Learning and Verbal Behaviour*, 20, 641-655.

D.G. Elliman & I.T. Lancaster, (1990), 'A Review of Segmentation and contextual analysis techniques for text recognition', *Pattern Recognition*, vol. 23, no. 3/4, 337-346.

Essex database, available from Department of Electronic Systems Engineering, University of Essex, UK.

L.J. Evett, C.J. Wells, F.G. Keenan, T.G. Rose & R.J. Whitrow, (1992), 'Using linguistic information to aid handwriting recognition', in S. Impedovo and J.C. Simon (eds), *From Pixels to Features III: Frontiers in Handwriting Recognition*, Elsevier: North-Holland.

L.J. Evett & G.J. Bellaby, (1994a), 'The Integration of Knowledge Sources for Word Recognition', *IEE European Workshop on Handwriting Analysis and Recognition*, Eds Prof M C Fairhurst and Dr A C Downton, Brussels , pp 16/1-16/8. 1994.

L.J. Evett, G.J. Bellaby, C. Wells, F. Keenan, T.G. Rose & R.J. Whitrow, (1994b), 'The Use of Linguistic Information in Script Recognition', *Joint MVA & IEE Technical Meeting on Handwritten Character and Script Recognition*, London, June 1993.

R. Farag, (1979), 'Word-level recognition of cursive script', *IEEE Transactions on Computers*, vol. 28, no. 2, February, 172-175.

I. Fischler & G.O. Goodman, (1978), 'Latency of associative activation in memory', *Journal of Experimental Psychology: Human Perception & Performance*, 4, 455-470.

B. Ford & W.P. Banks, (1977), 'Perceptual differences between reading handwritten and typed words', *Memory and Cognition*, 5, 630-635.

D.M. Ford & C.A. Higgins, (1990), 'A tree-based dictionary search technique and comparison with n-gram letter graph reduction', In R. Plamondon & C.G. Leedham (eds.), *Computer Processing of Handwriting*, World Scientific Publishing Co., 291-312.

G.D. Forney, Jr., (1973), 'The Viterbi algorithm', *Proceedings of the IEEE*, 61, no. 3, 268-278.

K.I. Forster, (1976), 'Accessing the mental lexicon', in R.J. Wales & E. Walker (eds.), *New Approaches to Language Mechanisms: A collection of psycholinguistic studies*, Amsterdam: North Holland.

J. Franke & E. Mandler, (1992), 'A comparison of two approaches for combining the votes of co-operating classifiers', *Proceedings of 11th IAPR International Conference on Pattern Recognition*', Vol. II, 611-614.

C. Frankish, P. Morgan & J. Noyes, (1994), 'Pen computing: some human factors issues', *IEE Colloquium on Handwriting and Pen-Based Input, March 1994*, Digest no: 1994/065.

H. Freeman, (1974), 'Computer processing of line-drawing images', *Computing Surveys*, vol. 6, no. 1, 57-97.

L.S. Frishkopf & L.D. Harmon, (1961), 'Machine reading of cursive script', In C. Cherry, (ed.), *Information Theory: 4th London Symposium*, London: Butterworths, 300-316.

L.S. Frishkopf, (1964), 'Automatic recognition of handwriting', U.S. Patent 3 133 266, May 1964.

G.O. Goodman, J.L McClelland & R.W. Gibbs Jr., (1981), 'The role of syntactic context in word recognition', *Memory and Recognition*, vol. 9, part 6, 580-586.

I. Guyon, L. Schomaker, R. Plamondon, M. Liberman & S. Janet, (1994), 'The UNIPEN project of data exchange and recognizer benchmarks', *Proceedings of the 12th ICPR*, October 9-13, Jerusalem.

R.N. Haber & R.M. Schindler, (1981), 'Errors in proof-reading: Evidence of syntactic control of letter processing?', *Journal of Experimental Psychology: Human Perception and Performance*, Part 7, 573-579.

C.A. Higgins & R.J. Whitrow, (1984), 'On-line cursive script recognition', *Proceedings of Interact '84, 1st IFIP Conference on Human Computer Interaction*, 139-143.

C.A. Higgins & D.M. Ford, (1991), 'Stylus driven interfaces - the electronic paper concept', *1st International Conference on Document Analysis and Recognition*, Saint-Malo, France, September, 853-862.

C.A. Higgins & P.E. Bramall, (1993), 'On-line Handwriting Recognition Based on the Human Reading Process', *Sixth International Conference on Handwriting and Drawing, ICOHD'93*, Paris, France, July, 111-113.

C.A. Higgins & P.E. Bramall, (1994), 'A system for cursive handwriting recognition based on human reading models', *IEE Colloquium on Handwriting and Pen-Based Input*, March 1994, Digest no: 1994/065, 7/1-7/7.

T.K. Ho, J.J. Hull & S.N. Srihari, (1992a), 'A word shape analysis approach to lexicon based word recognition', *Pattern Recognition Letters*, vol. 13, part 11, 821-826.

T.K. Ho, J.J. Hull & S.N. Srihari, (1992b), 'On multiple classifier systems for pattern recognition', *11th IAPR Int. Conf. on Pattern Recognition*, vol II, August 1992, 84.

D.H. Howes & R.L. Solomon, (1951), 'Visual duration as a function of word probability', *Journal of Experimental Psychology*, 41, 401-10.

Y.S. Huang & C.Y. Suen, (1993), 'An optimal method of combining multiple classifiers for unconstrained handwritten numeral recognition', *Proceedings of IWFHR-III, 3rd International Workshop on Frontiers in Handwriting Recognition.*

J.J. Hull, S.N. Srihari & R. Choudhari, (1983), 'An integrated algorithm for text recognition: comparison with a cascaded algorithm', *IEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-5, no. 4, July 1983, 384-395.

J.J. Hull & S.N. Srihari, (1986), "A computational approach to word shape recognition: Hypothesis generation and testing", *Proceedings IEEE-CS Conference on Computer Vision and Pattern Recognition*, 156-161.

G.W. Humphreys, L. Evett & D.E. Taylor, (1982), 'Automatic phonological priming in visual word recognition', *Memory and Cognition*, vol. 10(6), 576-590.

G.W. Humphreys, L. Evett & P.T. Quinlan, (1990), 'Orthographic processing in visual word identification', *Cognitive Psychology*, vol. 22, 517-560.

A.W. Inhoff & K. Rayner, (1986), 'Parafoveal word processing during eye fixations in reading: Effects of word frequency', *Perception and Psychophysics*, vol. 40, 431-439.

A.W. Inhoff, (1987), 'Parafoveal word perception during eye fixations in reading: Effects of visual salience and word structure', In M. Coltheart (ed.), *Attention and Performance XII: The Psychology of Reading*, London: Lawrence Erlbaum Associates, 403-418.

S. Johansson, (1980), 'The LOB corpus of British English texts: presentation and comments', *ALLC Journal*, vol. 1, no. 1.

J.C. Johnston, (1981), 'Understanding word perception: Clues from studying the word superiority effect', In O.J.L. Tzeng and H. Singer (eds.), *Perception of Print: Reading Research in Experimental Psychology*, Hillsdale, N.J.: Lawrence Erlbaum Associates.

N.F. Johnson, (1981), 'Integration processes in word recognition', In O.J.L. Tzeng and H. Singer (eds.), *Perception of Print: Reading Research in Experimental Psychology*, Hillsdale, N.J.: Lawrence Erlbaum Associates, 29-63.

M.A. Just, P. Carpenter & M.E.J. Masson, (1982), 'What eye fixations tell us about speed reading and skimming', (Technical Report), Pittsburgh: Carnegie-Mellon University

M.A. Just & P. Carpenter, (1987), *The Psychology of Reading and Language Comprehension*, Newton MA: Allyn and Bacon.

R.L. Kashyap & B.J. Oommen, (1981), 'An effective algorithm for string detection using generalized edit distances- I and II', *Information Sciences*, 23, 123-142 and 201-217.

R.L. Kashyap & B.J. Oommen, (1984), 'Spelling correction using probabilistic methods', *Pattern Recognition Letters*, 2(4), 147-154.

F.G. Keenan, (1993), 'Large vocabulary syntactic analysis for text recognition', Unpublished PhD Thesis, Nottingham Trent University.

F. Keenan & L.J. Evett, (1994), 'Applying syntactic information to text recognition', *Computational Linguistics for Speech and Handwriting Recognition*, April, AISB 1994 Workshop Series.

R.E. Kirk, (1990), *Statistics: An Introduction*, third edition, The William James Centennial Series, Fort Worth: Holt, Rinehart & Winston.

H. Kucera & W. Francis, (1967), *Computational analysis of present-day American English*, Providence, Rhode Island: Brown University Press.

E. Lecolinet & J.-P. Crettez, (1991), 'A grapheme-based segmentation technique for cursive script recognition', *1st International Conference on Document Analysis and Recognition*, Saint-Malo, France, September, 740-748.

E. Lecolinet, (1993), 'Cursive script recognition by backward matching', *Sixth International Conference on Handwriting and Drawing, ICOHD'93*, Paris, France, July, 89-91.

M. Lesk, (1987), 'Automatic sense discrimination using machine readable dictionaries: How to tell a pine cone from an ice cream cone', *Proc. 4th Conf. Assoc. for Computing Machinery*, 24-26.

V.I. Levenshtein, (1966), 'Binary codes capable of correcting deletions, insertions, and reversals', *Sov. Phys. Dokl.*, 10(8), 707-710.

E.D. Liddy & E.S. Yu, (1994), 'Text categorization for multiple users using semantic features from a machine readable dictionary', *ACM Transactions on Information Systems*, 12, 278-295.

S.D. Lima, (1987), 'Morphological analysis in sentence reading', *Journal of Memory and Language*, vol. 26, 84-99.

P.H. Lindsay & D.A. Norman, (1977), *Human Information Processing: An introduction to Psychology*, second edition, Orlando, Florida: Academic Press.

S. Madhvanath & S.V. Govindaraju, (1997), 'Contour-Based Image Preprocessing for Holistic Handwritten Word Recognition', *4th International Conference on Document Analysis and Recognition (ICDAR'97)*, Ulm, Germany, August 1997.

S. Madhvanath, V. Kripasundar & S.V. Govindaraju, (1997), 'Pruning Large Lexicons Using Generalised Word Shape Descriptors', *4th International Conference on Document Analysis and Recognition, (ICDAR'97)*, Ulm, Germany, August 1997.

S. Madhvanath & S.V. Govindaraju, (1998), 'Perceptual Features for Off-Line handwritten Word Recognition- A Framework for Heuristic Prediction, Matching and Representation', *Syntactic and Statistical Pattern Recognition*, Sydney, Australia, Aug. 1998.

M. Malburg, (1997), 'Comparative Evaluation of Techniques for Word Recognition Improvement by Incorporation of Syntactic Information', (1997), *International Conference on Document Analysis and Recognition (ICDAR'97)*, Ulm, Germany, August 199 7.

E. Mandler & J. Schurmann, (1988), 'Combining the classification results of independent classifiers based on the Dempster/Shafer theory of evidence', *Pattern Recognition in Practice*, Amsterdam, 381-394.

C. Manso de Zuniga, G.W. Humphreys & L.J. Evett, (1991), 'Additive and interactive effects of repetition, degradation and word frequency in the reading of handwriting', In D. Besner & G.W. Humphreys (eds.), *Basic Processes in Reading: Visual Word Recognition*, 10-33.

M. Mason, (1975), 'Reading ability and letter search time: Effects of orthographic structure defined by single-letter positional frequency', *Journal of Experimental Psychology: General*, 104, 146-166.

D.W. Massaro, R.L. Venezky & G.A. Taylor, (1979), 'Orthographic regularity, positional frequency, and visual processing of letter strings', *Journal of Experimental Psychology: General*, 108, 107-124.

J.L. McClelland, (1977), 'Letter and configuration information in word identification', *Journal of Verbal Learning and Verbal Behaviour*, 16, 137-150.

J.L. McClelland & D.E. Rumelhart, (1981), 'An interactive activation model of context effects in letter perception: Part 1. An account of basic findings', *Psychological Review*, vol. 88, no. 5, 375-407.

J.L. McClelland, (1986), 'The programmable blackboard model of reading', In J.L. McClelland & D.E. Rumelhart, (eds.), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, vol. 2, Cambridge, Mass.: MIT Press.

J.L. McClelland, M. St. John & R. Taraban, (1992), 'Sentence comprehension: A parallel distributed processing approach', In L. Burkholder (ed.), *Philosophy and the Computer*, Boulder, Colorado: Westview Press, 34-56.

G.W. McConkie & D. Zola, (1979), 'Is visual information integrated across successive fixations in reading?', *Perception and Psychophysics*, vol. 25, 221-224.

G.W. McConkie, D. Zola, H.E. Blanchard & G.S. Wolverton, (1982), 'Perceiving words during reading. Lack of facilitation from prior peripheral exposure', *Perception and Psychophysics*, vol. 32, 271-281.

D.E. Meyer, R.W. Schvaneveldt & M.G. Ruddy, 1975, 'Loci of contextual effects on visual word-recognition', In P.M.A. Rabbitt & S. Dornic, (eds.), *Attention and Performance V*, London: Academic Press.

G. Miller & S. Isard, (1963), Some perceptual consequences of linguistic rules, *Journal of Verbal Learning and Verbal Behaviour*, 2. 217-228.

G.M. Miller, (1971), 'Real-time classification of handwritten script words', *Proceedings IFIP Congress on Information Processing 1971*, Amsterdam: North Holland, 1972.

S. Monsell, (1991), 'The nature and locus of word frequency effects in reading', In D. Besner & G.W. Humphreys (eds.), *Basic Processes in Reading: Visual Word Recognition*, 148-197.

J. Morton, (1969), 'Interaction of information in word recognition', *Psychological Review*, 76, 165-178.

M.C. Mozer, (1987), 'Early parallel processing in reading: A connectionist approach', In M. Coltheart (ed.), *Attention and Performance XII: The Psychology of Reading*, London: Lawrence Erlbaum Associates, 83-104.

C. Nadal & C.Y. Suen, (1993), 'Applying human knowledge to improve machine recognition of confusing handwritten numerals', *Pattern Recognition*, vol. 26, no. 3, 381-389.

S.C. Oh, J.Y. Ha, & J.H. Kim, (1995), 'Context-Dependent Search in Interconnected Hidden Markov Model for Unconstrained Handwriting Recognition', *Pattern Recognition,* (28), No.11, November 1995, 1693-1704.

J.K. O'Regan, (1979), 'Saccade size control in reading: Evidence for the linguistic control hypothesis', *Perception and Psychophysics*, vol. 25, no. 6, 501-509.

J.K. O'Regan, (1981), '"Convenient viewing position" hypothesis', In D.F. Fisher, R.A. Monty & J.W. Senders, (eds.), *Eye movements: Cognition and visual perception*, Hillsdale, N.J.: Lawrence Erlbaum Associates.

R.K. Powalka, N. Sherkat, L.J. Evett & R.J. Whitrow, (1993), 'Multiple word segmentation with interactive look-up for cursive script recognition', *Second*

International Conference on Document Analysis and Recognition (ICDAR'93), Tsukuba Science City, Japan, October, 196-199.

R.K. Powalka, N. Sherkat & R.J. Whitrow, (1994), 'Feature extraction: on the importance of zoning information in cursive script recognition', In S. Impedovo (ed.), *Progress in Image Analysis and Processing III, World Scientific*, 342-349.

R.K. Powalka, (1995), 'Algorithm toolbox for on-line cursive script', Unpublished PhD Thesis, Nottingham Trent University.

J.E. Ranklin, (1977), 'Psycholinguistic cues in peripheral vision during reading', Unpublished Ph.D thesis, University of London.

K. Rayner, G.W. McConkie & D. Zola, (1980), 'Integrating information across eye movements', *Cognitive Psychology*, vol. 12, 206-226.

K. Rayner & A. Pollatsek, (1987), 'Eye movements in reading: A tutorial review', In M. Coltheart (ed.), *Attention and Performance XII: The Psychology of Reading*, London: Lawrence Erlbaum Associates, 327-362.

K. Rayner, A.D. Well, A. Pollatsek, & J.H. Bertera, (1982), 'The availability of useful information to the right of fixation in reading', *Perception and Psychophysics*, vol. 31, 537-550.

K. Rayner & D.A. Balota, (1989), 'Parafoveal preview effects and lexical access during eye fixations in reading', In W. Marslen-Wilson, Lexical *Representation and Process*, Cambridge, Mass.: Bradford Book/MIT Press, 261-290.

R. Reddy, (1975), (ed.), *Speech Recognition: Invited papers presented at the IEEE symposium*, New York: Academic Press.

E.M. Riseman & R.W. Ehrich, (1971), 'Contextual word recognition using binary diagrams', *IEEE Transactions on Computers*, vol. 20, non. 4, 397-403.

T.G. Rose & L.J. Evett, (1992), 'A large vocabulary semantic analyser for handwriting recognition', *AISB Quarterly*, 80, 34-39.

T.G. Rose & L.J. Evett, (1993), 'Semantic analysis for large vocabulary cursive script recognition', *Proceedings of the Second International Conference on Document Analysis and Recognition*, Tsukuba Science City, Japan.

T.G. Rose, (1994), 'Large vocabulary semantic analysis for text recognition', Unpublished PhD Thesis, Nottingham Trent University.

D.E. Rumelhart & J.L. McClelland, (1982), 'An interactive activation model of context effects in letter perception: Part 2. The contextual enhancement effect and some tests and extensions of the model', *Psychological Review*, 89, 60-94.

G. Sampson, (1987), 'Evidence against the "grammatical"/"ungrammatical" distinction', In W. Meijs, (ed.), *Corpus Linguistics and Beyond: Proceedings of the 7th International Conference on English Language Research on Computerized Corpora*, 219-226.

E. Sapir, (1921), *Language: A Study of Speech*, Harcourt: Brace & World.

L. Schomaker, (1994), 'User-interface aspects in recognizing connected-cursive handwriting', *IEE Colloquium on Handwriting and Pen-Based Input*, March 1994, Digest no: 1994/065,

C.E. Shannon, (1948), A mathematical theory of communication, *The Bell System Technical Journal*, 27, 379-423.

C.E. Shannon, (1950), 'Prediction and entropy of printed English', *Bell System Technical Journal*, vol. 30, 50-64.

R. Shinghal & G.T. Toussaint, (1979), 'A bottom-up and top-down approach to using context in text recognition', *International Journal of Man-Machine Studies*, vol. 11, 201-212.

J.-C. Simon, (1992), 'Off-line cursive script recognition', *Proceedings of the IEEE*, vol. 80, no. 7, 1150-1161.

S.N. Srihari & R.M. Bozinovic, (1982), 'A string correction algorithm for cursive script recognition', *Proceedings of the 6th International Conference on Pattern Recognition*, (Munich W. Germany), IEEE, New York, 232-234.

S.N. Srihari, J.J. Hull & R. Choudhari, (1983), 'Integrating diverse knowledge sources in text recognition', *ACM Transactions on Office Information Systems*, vol. 1, part 1, 68-87.

S.N. Srihari & R.M. Bozinovic, (1987), 'A multi-level perception approach to reading cursive script', *Artificial Intelligence*, 33, 217-255.

R. Srihari & C.M. Baltus, (1993), 'Incorporating syntactic constraints in recognizing handwritten sentences', *Proc. of the Int. Joint Conf. on Artificial Intelligence, IJCAI-93*, 1262-126, Chambery, France, August 1993.

S. N. Srihari & E.J. Keubert, (1997), 'Integration of hand-written address interpretation technology into the United States postal service remote computer reader system, *4th International Conference on Document Analysis and Recognition (ICDAR'97)*, Ulm, Germany, August 1997.

C.Y. Suen, (1983), 'Handwriting generation, perception and recognition', *Acta Psychologica*, vol. 54, 295-312.

C.Y. Suen, C. Nadal, T.A. Mai, R. Legault, L. Lam, (1990), 'Recognition of totally unconstrained handwritten numerals based on the concept of multiple experts', *Proceedings of the Conference on Frontiers in Handwriting Recognition*, Concordia University, Montreal Canada, 1990, 131-144.

M.K. Tannenhaus, H.P. Flanigan & M.S. Seidenberg, (1980), 'Orthographic and phonological activation in auditory and visual word recognition', *Memory and Cognition*, vol. 8, part 6, 513-520.

C.C. Tappert, (1982), 'Cursive script recognition by elastic matching', *I.B.M. Journal of Research and Development*, volume 26, no. 6, Nov. 1982, 765-771.

I. Taylor & M.M. Taylor, (1983), *The Psychology of Reading*, New York: Academic Press.

M.C. Thompson & D.W. Massaro, (1973), 'Visual information and redundancy in writing', *Journal of Experimental Psychology*, 98, 49-54.

E. Tulving & C. Gold, (1963), 'Stimulus information and contextual information as determinants of tachistoscopic recognition of words', *Journal of Experimental Psychology*, 66, 319-327.

A.J. Viterbi, (1967), 'Error bounds for convolutional codes and an asymptoticaly optimum decoding algorithm', *IEEE Transactions on Information Theory*, IT-13, April, 269-269.

C.J. Wells, L.J. Evett, P.E. Whitby & R.J. Whitrow, (1990), 'Fast dictionary look-up for contextual word recognition', *Pattern Recognition*, vol. 23, no. 5, 501-508.

C.J. Wells, L.J. Evett & R.J. Whitrow, (1991), 'Word look-up for script recognition- choosing a candidate', *Proceedings of the 1st International Conference on Document Analysis and Recognition, (ICDAR-91)*, 620-628.

C.J. Wells, (1992), 'The use of orthographic and lexical information for handwriting recognition', Unpublished PhD Thesis, Nottingham Trent University.

M. Wilson, (1984), 'The composition of the mental lexicon', Unpublished PhD Thesis, Cambridge University.

P.T. Wright, (1989), 'Algorithms for the recognition of handwriting in real-time', Unpublished PhD Thesis, Nottingham Trent University.

C.G. Wolf, (1990), 'Understanding handwriting recognition from the user's perspective', *Proceeding of the 34th Annual Meeting of the Human Factors Society*, 249-253.

L. Xu, A. Krzyzak & C.Y. Suen, (1992), 'Methods of combining classifiers and their applications to handwriting recognition', *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 22, no. 3, May/June 1992, 418-435.

Appendix A: Introduction

26 word data set

claim	share	goes	cover
baby	sides	spoke	laws
check	enjoy	equal	exact
older	takes	trees	royal
hell	fell	pass	crazy
safe	names	noted	
sign	dance	usual	

10 word two letter data set

ci decide

lo glow

cl cycle

hi chief

li climb

ri drift

rn burnt

un tune

ll rally

ui fruit

Appendix B: Analysis of the Lexicon by the Seven Cues

The lexicon is partitioned by different cues to varying degrees. Seven figures are given in this appendix, each showing how the lexicon is partitioned by one of the seven cues (see section 4.2).

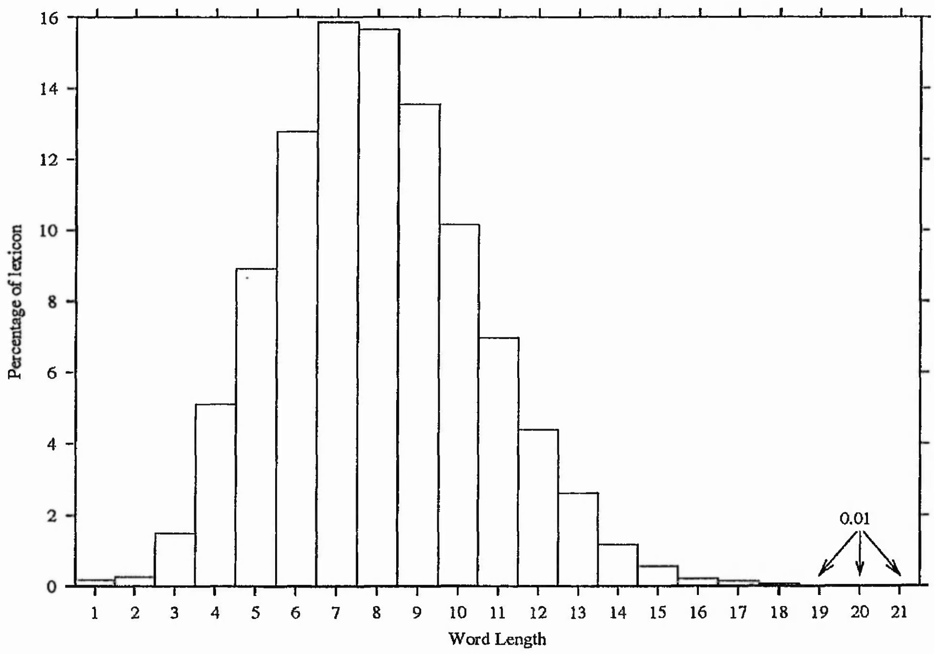


Figure B-1: Analysis of the lexicon by word length

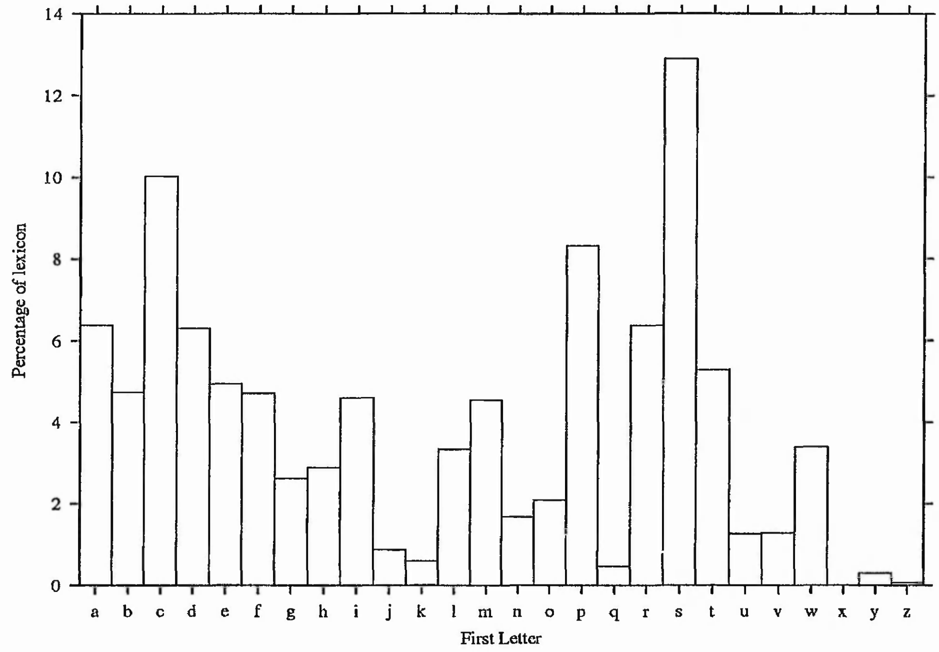


Figure B-2: Analysis of the lexicon by first letter

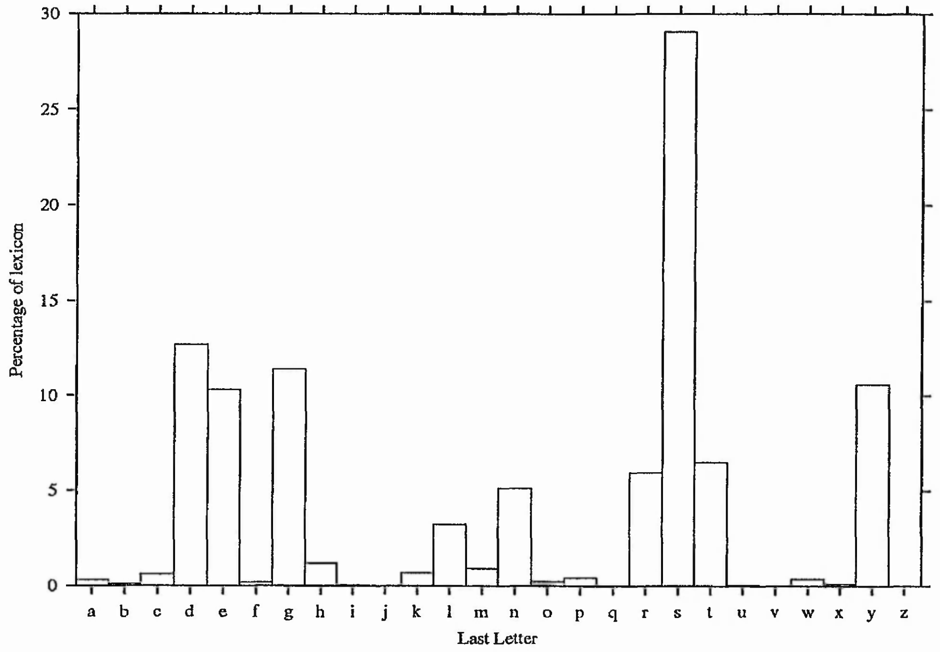


Figure B-3: Analysis of the lexicon by last letter

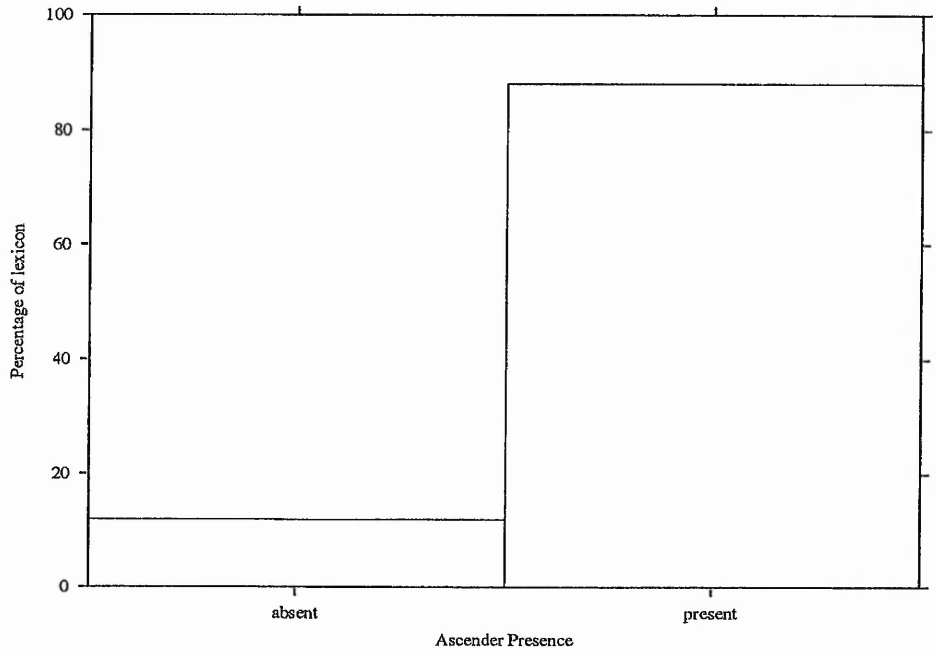


Figure B-4: Analysis of the lexicon by ascender presence

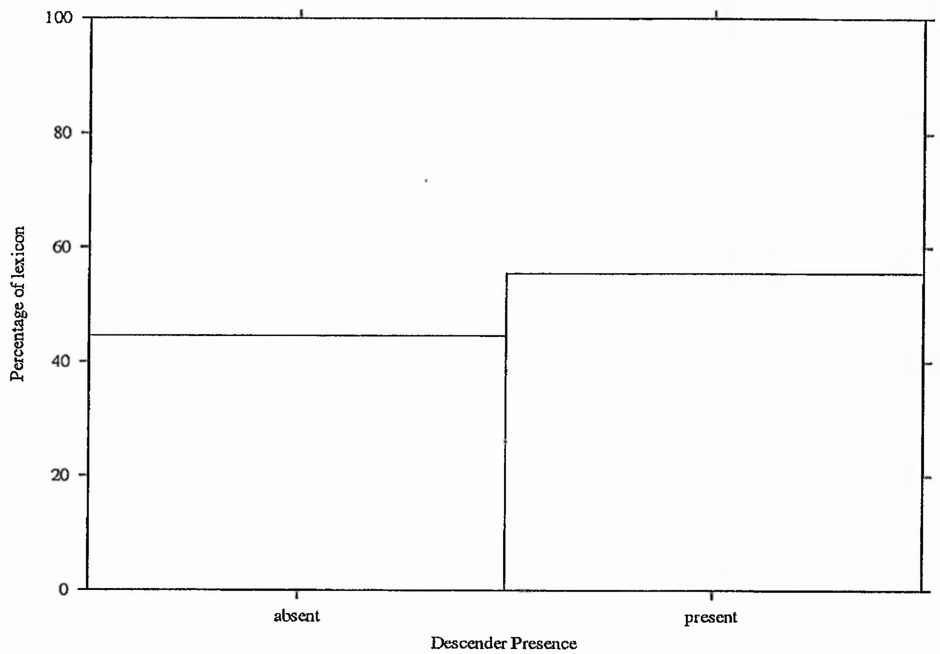


Figure B-5: Analysis of the lexicon by descender presence

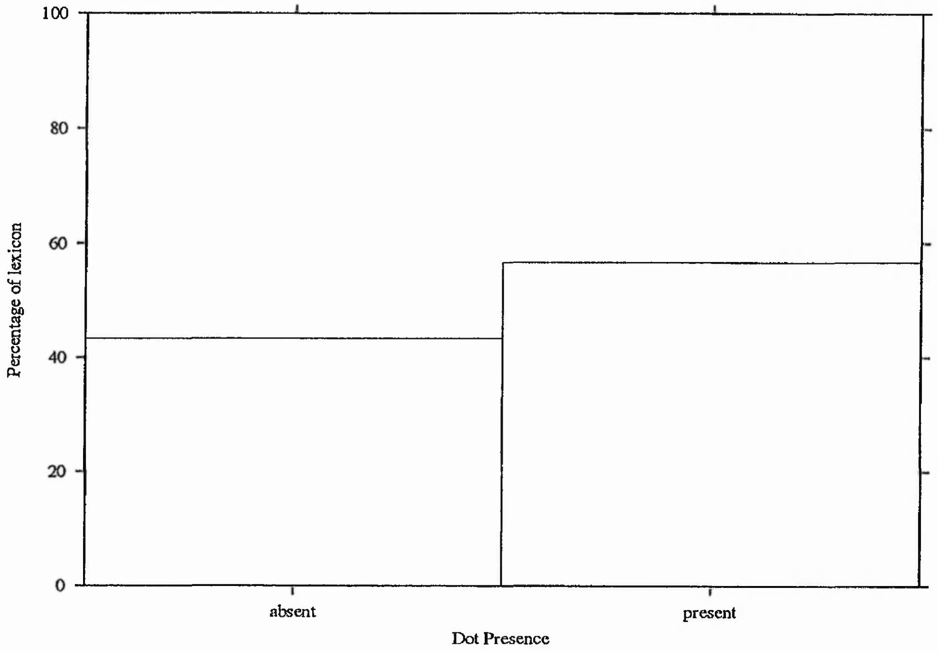


Figure B-6: Analysis of the lexicon by dot presence

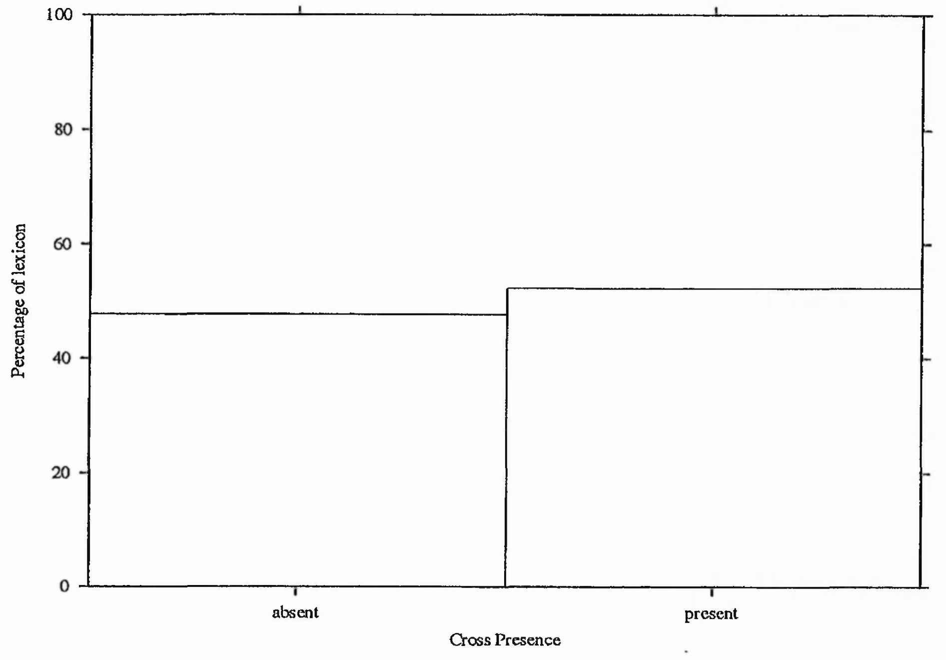


Figure B-7: Analysis of the lexicon by cross presence

Appendix C: Degree of Resemblance Between Candidates and Target

The following two figures show the degree of resemblance of output from the pattern recognizer with the target word (see section 4.3.4). Results are given for the partial data set. The number of cues in each candidate word that were identical to the cues of the target word is shown.

2% of the partial data set had a list size of 100. In the partial data set the average number of candidates generated by the pattern recognizer was 19. The lowest confidence score was 22. 1.7% of the candidates had a confidence score of 22. 17% of the word lists had one or more candidates with a confidence score of 22. The average confidence score of candidates in the partial data set was 50.

Figure C-1 shows the degree of resemblance between the candidates and their target word by the ranked position of the candidate.

Figure C-2 shows the degree of resemblance between the candidates and their target word by confidence score of the candidate.

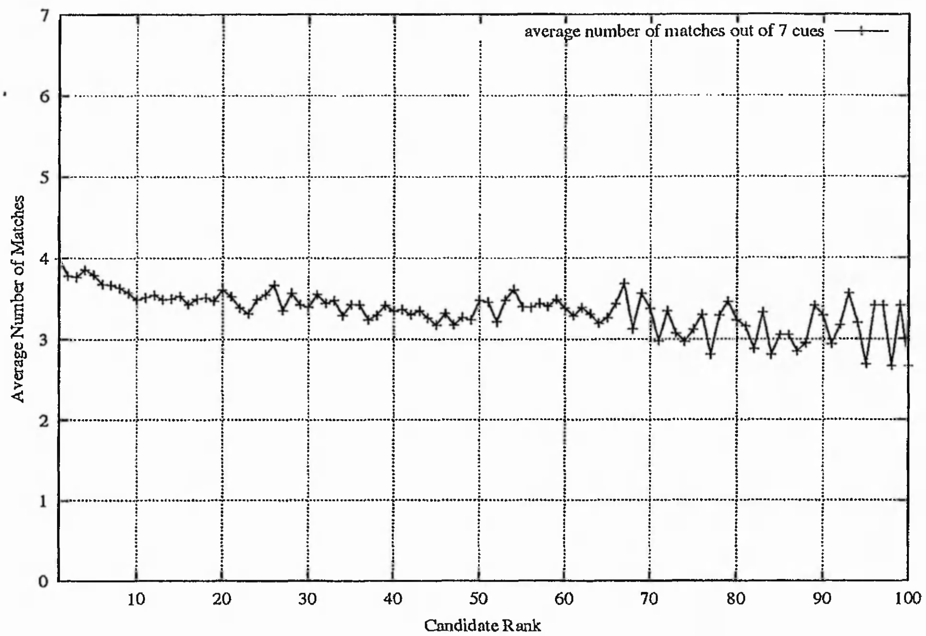


Figure C-1: Degree of resemblance to the target word by rank

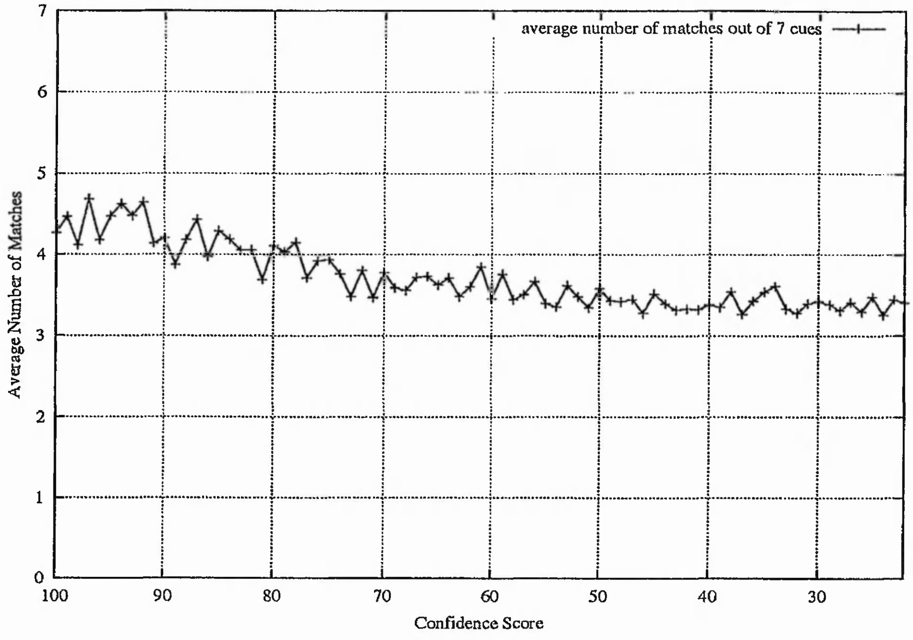


Figure C-2: Degree of resemblance to the target word by score

Appendix D: 200 Word Data Set

a
address
advantage
adventure
alternative
alternatives
am
amazing
an
analysis
and
answer
architecture
are
assembly
available
billion
brandy
can
champagne
cheque
communication
complex
computer
computing
cursive
demonstration
design
development
different
difficult
does
eight
eighteen
eighty
electronic
eleven
example
exhibition
expedition
experiment
false
fifteen
fifty
figure
find
five
flexible
for
form
forms
formula
forty
four
fourteen
gesture
good
grammar
guidance
handwriting
has
have
he
hello
honest
human
hundred
i
important
impulse
in
interest
interesting
interface
international
interpretation
introduction
invitation
is
it
journal
just
knowledge
language
legend
magazine
mathematical
mathematics
me
memory
million
minimum
modern
necessary
never
new
nine
nineteen
ninety
no
not
nothing
number
observation
of
off
office
on
one
opportunity
oxygen
painting
paper
particular
particularly
people
perhaps
period
phenomenon
place
point
possible
power
present

probably
problem
professional
professor
program
programme
psychology
question
quite
read
really
recognition
recognize
recognizer
representatives
requirement
responsibilities
responsibility
scientific
script
second
seven
seventeen
seventy
she
six
sixteen
sixty
small
software
something
source
special
story
studies
subject
successful
suggestion
sunlight
system
ten
terrible
the
they
thirteen
thirty
this
thousand
three
time
to
twelve
twenty
two
under
understanding
until
use
used
usually
vehicle
very
view
we
weakness
what
where
which
whisky
word
words
world
yes
you
your
zero

Appendix E: Comparison of the 200 Word Data Set and the Common Lexicon

Figure E-1 through to Figure E-8 show comparisons between the distribution of various relevant factors in the 200 word data set and in the common lexicon (see section 4.3.6.2).

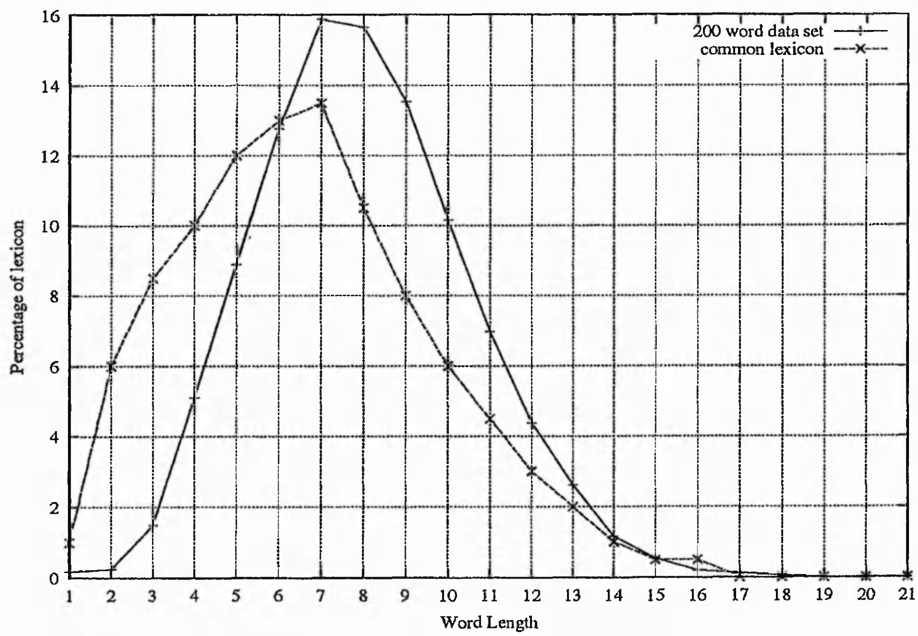


Figure E-1: Comparison of the distribution of word length in the 200 word data set and in the common lexicon

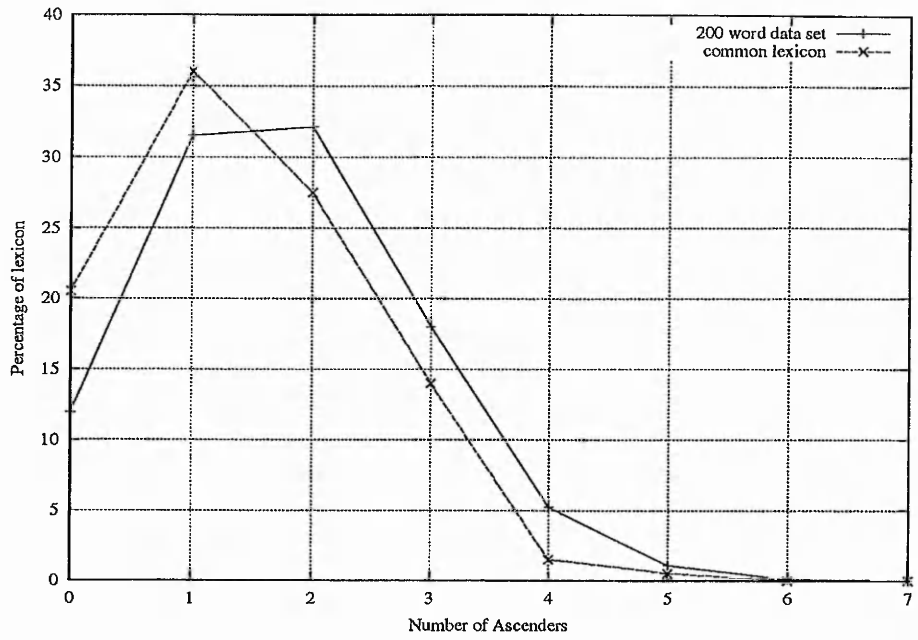


Figure E-2: Comparison of the distribution of number of ascenders in the 200 word data set and in the common lexicon

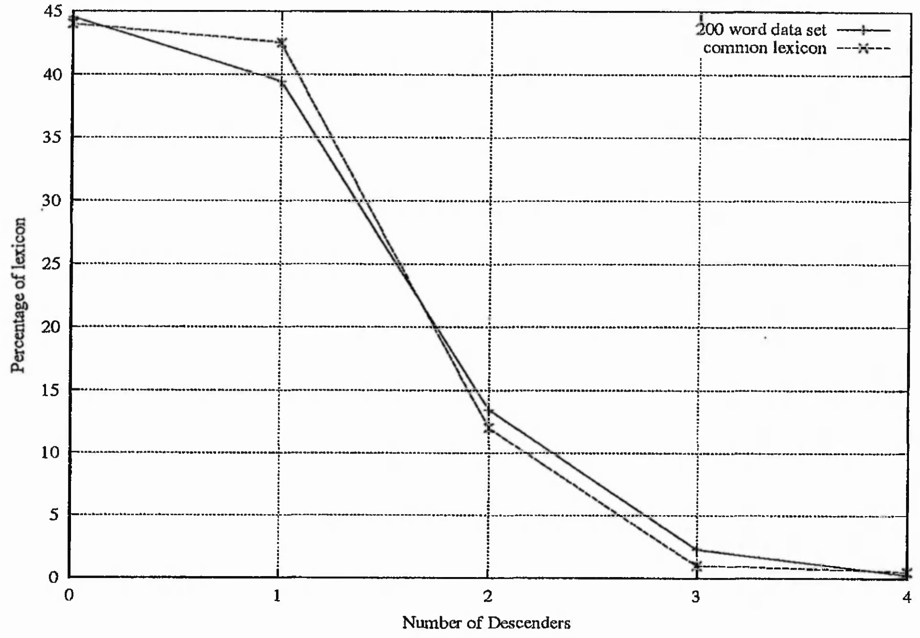


Figure E-3: Comparison of the distribution of number of descenders in the 200 word data set and in the common lexicon

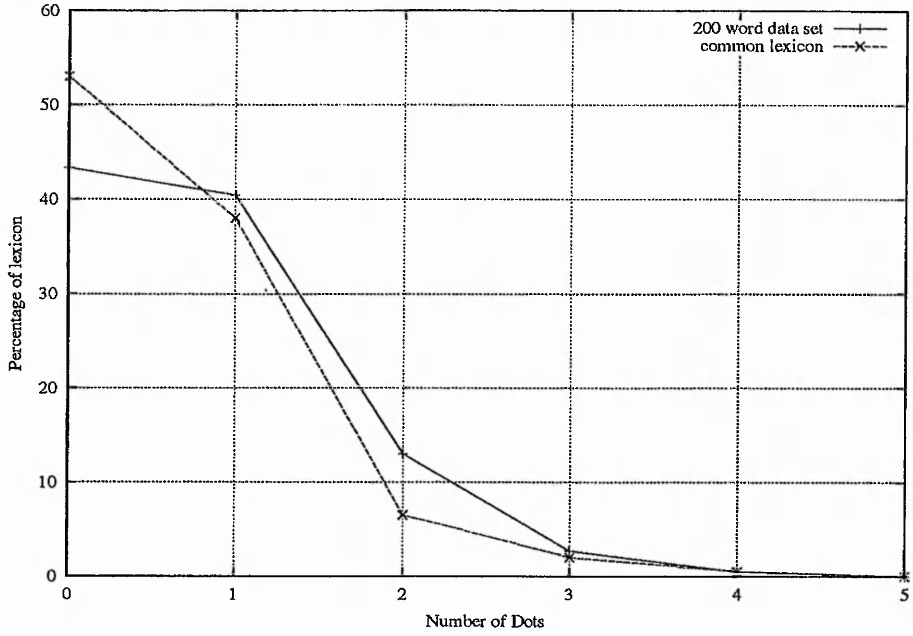


Figure E-4: Comparison of the distribution of number of dots in the 200 word data set and in the common lexicon

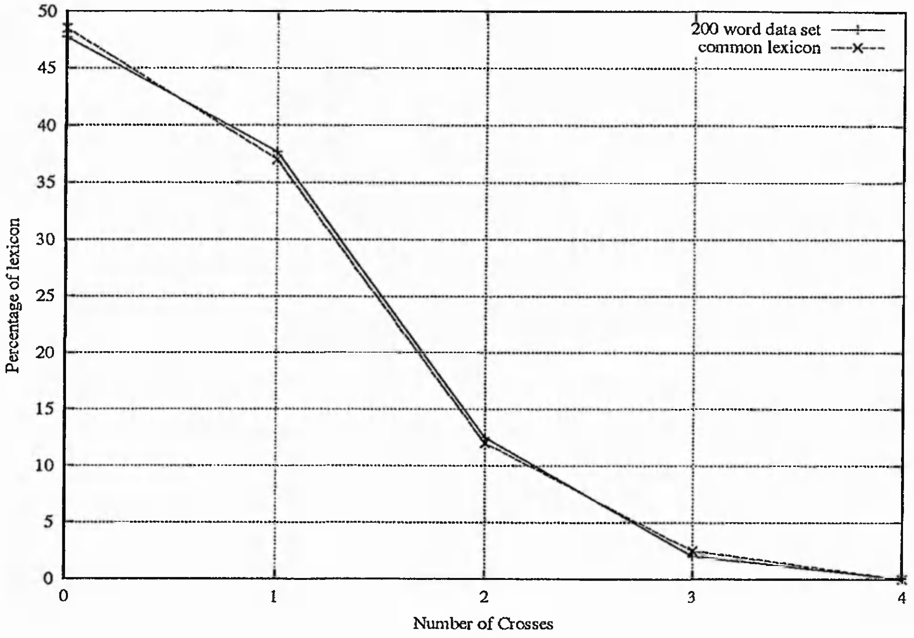


Figure E-5: Comparison of the distribution of number of crosses in the 200 word data set and in the common lexicon

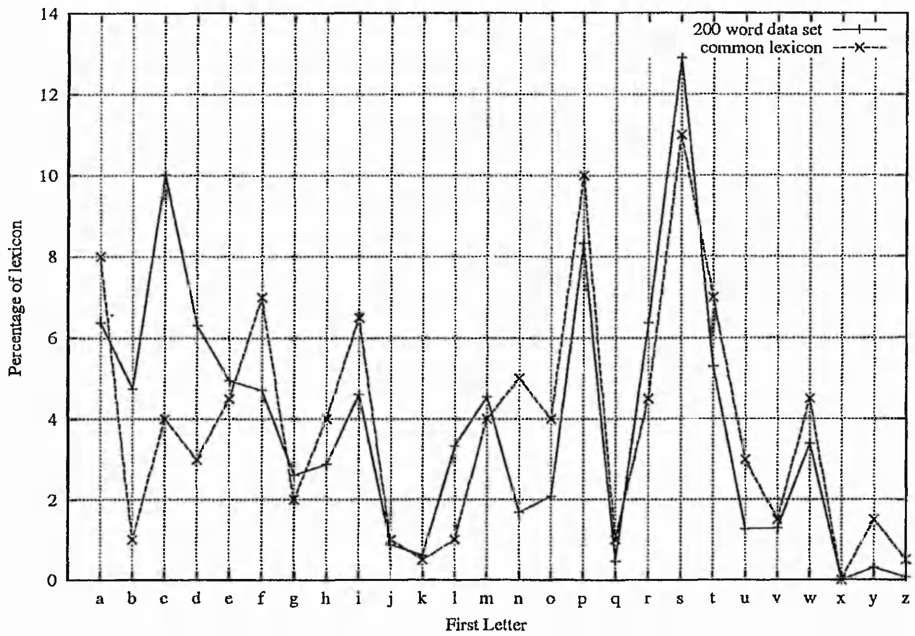


Figure E-6: Comparison of the distribution of first letters in the 200 word data set and in the common lexicon

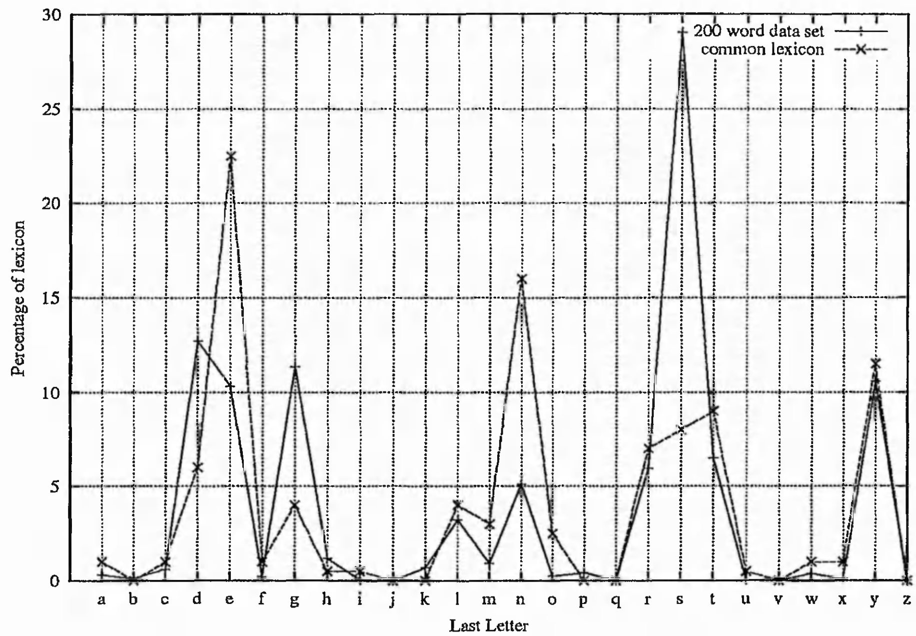


Figure E-7: Comparison of the distribution of last letters in the 200 word data set and in the common lexicon

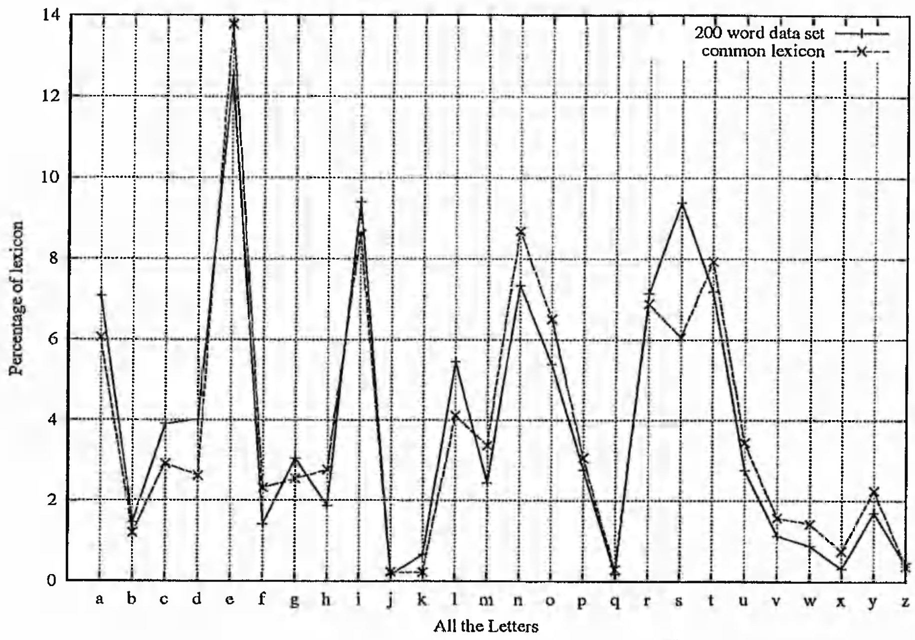


Figure E-8: Comparison of the distribution of all of the letters in the 200 word data set and in the common lexicon

Table E-1 shows indices of correlation between the distribution of relevant factors in the 200 word data set and the common lexicon. Two indices of correlation are given: the Pearson product-moment correlation coefficient and the Searman rank correlation. The former measures the strength of the relationship between two quantitative variables, whilst the latter measures the degree of agreement between two sets of ranks.

It is not useful to calculate correlation coefficients for the cues presence or absence of ascenders, descenders, i-dots and j-dots, and t-crosses and f-crosses since these are binary cues and therefore the number of paired scores would be 2, i.e. present and absent. For all of these cues, the number of times a property is present is used to calculate the correlation instead.

It is also not useful to calculate the correlation coefficient for the unadjusted word frequency of the words because 99% of the words in the 200 word data set occur with a distribution score of 1. Word frequency has therefore been banded in order to ensure that distribution scores are high enough for a useful comparison to be made between the 200 word data set and the common lexicon.

cue	Pearson product-moment correlation coefficient	Spearman rank correlation
word length	r = 0.83, t = 6.45, df = 19	r = 0.90, t = 9.06, df = 19
number of ascenders	r = 0.95, t = 7.26, df = 6	r = 0.95, t = 7.16, df = 6
number of descenders	r = 1.00, t = 23.41, df = 3	r = 1.00, t = 23.41, df = 3
number of dots	r = 0.98, t = 9.36, df = 4	r = 1.00, t= 9.36, df = 4
number of crosses	r = 1.00, t = 61.70, df = 3	r = 1.00, t = 61.70, df = 3
first letter	r = 0.77, t = 5.85, df = 24	r = 0.75, t = 5.58, df = 24
last letter	r = 0.56, t = 3.35, df = 24	r = 0.81, t = 6.82, df = 24
all the letters	r = 0.95, t = 15.28, df = 24	r = 0.95, t = 14.54, df = 24
word frequency	r = 0.70, t = 11.12, df = 127	r = 0.57, t = 7.73, df = 127

Table E-1: Indices of correlation between the distribution of various relevant factors within the 200 word data set and the common lexicon

It can be seen from the above table that all of the distributions are comparable.

Appendix F: Cue Derivation- Initial Results

The following eight tables show results for cue derivation. Results are for the partial data set. The tables are divided, firstly, by the method used to limit the number of candidates (by rank or by confidence score), and, secondly, by whether the calculations were weighted by the confidence score of the candidates, or whether the calculation were unweighted (see section 4.3.6.3).

In Table F-1 to Table F-4, the results are given by the threshold used to limit the number of candidates. The combination of alternatives that produced the most accurate results for a given cue has been used for all of the other methods.

threshold	length	ascender	descender	dot	cross	first	last	average
1	33.4%	80.4%	71.5%	74.0%	66.1%	35.1%	40.2%	57.2%
2	34.4%	80.4%	71.2%	78.1%	66.1%	35.1%	40.2%	57.9%
3	37.7%	82.0%	70.7%	74.1%	67.5%	34.8%	41.2%	58.3%
4	37.4%	82.2%	71.0%	77.4%	67.4%	35.4%	42.0%	59.0%
5	37.2%	81.1%	70.8%	74.3%	67.9%	37.2%	41.5%	58.6%
6	37.6%	80.9%	70.8%	76.1%	67.5%	37.4%	41.7%	58.9%
7	37.2%	80.7%	71.2%	75.0%	67.5%	38.2%	41.8%	58.8%
8	37.2%	80.9%	72.0%	75.6%	68.0%	38.4%	42.0%	59.2%
9	37.1%	80.9%	71.3%	74.5%	67.5%	37.9%	41.8%	58.7%
10	37.7%	80.9%	71.2%	75.1%	68.4%	37.6%	41.8%	59.0%
11	37.6%	80.6%	70.7%	74.0%	68.7%	37.6%	41.8%	58.7%
12	37.9%	80.6%	70.8%	74.5%	68.4%	37.2%	41.7%	58.7%
13	38.2%	80.4%	71.0%	74.1%	67.7%	36.6%	41.2%	58.5%
14	38.4%	80.6%	70.5%	74.6%	68.4%	36.1%	41.2%	58.5%
15	38.1%	80.7%	70.7%	74.5%	68.0%	36.2%	41.7%	58.6%

threshold	length	ascender	descender	dot	cross	first	last	average
16	37.4%	81.1%	71.2%	74.6%	68.4%	36.1%	41.7%	58.6%
17	37.6%	80.9%	71.3%	74.5%	68.5%	36.1%	41.5%	58.6%
18	37.4%	80.9%	71.3%	74.5%	68.0%	36.4%	41.5%	58.6%
19	36.7%	80.9%	71.2%	74.1%	67.9%	36.2%	41.4%	58.3%
20	36.7%	80.7%	71.2%	74.5%	67.9%	36.2%	41.4%	58.4%
21	36.9%	80.4%	70.8%	74.1%	67.9%	35.9%	41.0%	58.2%
22	37.2%	80.6%	70.8%	74.6%	67.7%	35.7%	41.2%	58.3%
23	37.2%	80.4%	71.0%	74.5%	67.7%	35.6%	41.2%	58.2%
24	37.1%	80.7%	70.7%	74.5%	68.0%	35.7%	41.0%	58.2%
25	37.2%	80.6%	70.7%	74.5%	68.2%	35.9%	41.5%	58.4%
26	37.2%	80.7%	70.8%	74.6%	68.0%	36.1%	41.4%	58.4%
27	37.4%	80.6%	70.7%	74.5%	67.9%	35.7%	41.8%	58.4%
28	37.2%	80.6%	71.0%	74.5%	68.4%	35.7%	41.8%	58.5%
29	37.1%	80.6%	71.0%	74.5%	68.0%	35.7%	41.7%	58.4%
30	37.1%	80.6%	70.8%	74.3%	68.2%	35.7%	41.5%	58.3%
31	37.1%	80.4%	70.7%	74.5%	68.2%	35.7%	41.7%	58.3%
32	36.9%	80.6%	70.7%	74.6%	68.4%	35.6%	41.5%	58.3%
33	36.9%	80.6%	70.5%	74.6%	68.2%	35.6%	41.5%	58.3%
34	37.1%	80.4%	70.5%	74.6%	68.4%	35.6%	41.8%	58.3%
35	37.1%	80.2%	70.3%	74.6%	68.2%	35.6%	42.0%	58.3%
36	36.9%	80.4%	70.2%	74.6%	68.4%	35.6%	42.0%	58.3%
37	36.7%	80.4%	70.2%	74.8%	68.4%	35.6%	42.0%	58.3%
38	36.7%	80.6%	70.3%	74.8%	68.4%	35.6%	42.2%	58.4%
39	36.7%	80.6%	70.2%	74.8%	68.2%	35.6%	42.3%	58.3%
40	36.9%	80.6%	70.2%	74.8%	68.2%	35.6%	42.5%	58.4%
41	36.9%	80.6%	69.9%	74.8%	68.4%	35.7%	42.5%	58.4%
42	36.9%	80.6%	69.7%	74.8%	68.5%	35.7%	42.3%	58.4%
43	36.9%	80.6%	69.7%	74.8%	68.4%	35.7%	42.2%	58.3%
44	36.9%	80.6%	69.7%	74.8%	68.4%	35.9%	41.8%	58.3%
45	37.1%	80.4%	69.7%	75.0%	68.4%	36.1%	42.2%	58.4%
46	37.1%	80.6%	69.7%	75.0%	68.4%	35.7%	42.3%	58.4%
47	37.1%	80.6%	69.7%	75.0%	68.5%	35.6%	42.3%	58.4%
48	36.9%	80.4%	69.7%	75.0%	68.5%	35.7%	42.2%	58.3%

threshold	length	ascender	descender	dot	cross	first	last	average
49	36.7%	80.4%	69.7%	75.0%	68.4%	35.7%	42.0%	58.3%
50	36.7%	80.4%	69.7%	75.0%	68.4%	35.7%	42.0%	58.3%
51	36.9%	80.4%	69.7%	74.8%	68.5%	35.6%	42.3%	58.3%
52	36.9%	80.4%	69.7%	74.8%	68.5%	35.7%	42.3%	58.3%
53	36.6%	80.4%	69.7%	74.8%	68.5%	35.7%	42.3%	58.3%
54	36.6%	80.4%	69.7%	74.8%	68.5%	35.6%	42.5%	58.3%
55	36.6%	80.2%	69.7%	74.8%	68.4%	35.6%	42.3%	58.2%
56	36.6%	80.2%	69.7%	74.8%	68.5%	35.4%	42.3%	58.2%
57	36.4%	80.2%	69.7%	74.8%	68.4%	35.4%	42.3%	58.2%
58	36.6%	80.2%	69.7%	74.8%	68.4%	35.3%	42.3%	58.2%
59	36.7%	80.2%	69.7%	74.8%	68.4%	35.3%	42.3%	58.2%
60	36.7%	80.2%	69.7%	74.8%	68.4%	35.3%	42.3%	58.2%
61	36.6%	80.2%	69.7%	74.8%	68.4%	35.3%	42.3%	58.2%
62	36.6%	80.2%	69.7%	74.8%	68.2%	35.3%	42.3%	58.2%
63	36.6%	80.2%	69.7%	74.8%	68.2%	35.1%	42.5%	58.2%
64	36.7%	80.2%	69.7%	74.8%	68.2%	35.1%	42.5%	58.2%
65	36.7%	80.2%	69.9%	74.8%	68.2%	35.1%	42.5%	58.2%
66	36.7%	80.2%	69.9%	74.8%	68.2%	35.1%	42.7%	58.2%
67	36.7%	80.2%	69.9%	74.8%	68.2%	35.4%	42.7%	58.3%
68	36.7%	80.2%	69.9%	74.8%	68.2%	35.4%	42.7%	58.3%
69	36.7%	80.2%	69.9%	74.8%	68.2%	35.4%	42.7%	58.3%
70	36.9%	80.2%	69.9%	74.8%	68.2%	35.4%	42.7%	58.3%
71	36.9%	80.2%	69.9%	74.8%	68.0%	35.1%	42.5%	58.2%
72	36.9%	80.2%	69.9%	75.0%	68.0%	35.1%	42.5%	58.2%
73	36.9%	80.2%	69.9%	75.0%	68.2%	35.1%	42.5%	58.2%
74	36.9%	80.2%	69.9%	75.0%	68.2%	35.1%	42.5%	58.2%
75	36.9%	80.2%	69.9%	75.0%	68.2%	35.1%	42.5%	58.2%
76	36.9%	80.2%	69.9%	75.0%	68.2%	35.1%	42.5%	58.2%
77	36.9%	80.2%	69.9%	75.0%	68.2%	35.1%	42.5%	58.2%
78	36.9%	80.2%	69.9%	75.0%	68.2%	35.1%	42.5%	58.2%
79	36.9%	80.2%	69.9%	75.0%	68.2%	35.1%	42.7%	58.3%
80	36.9%	80.2%	69.9%	75.0%	68.2%	35.1%	42.7%	58.3%
81	36.9%	80.2%	69.9%	75.0%	68.2%	35.1%	42.7%	58.3%

threshold	length	ascender	descender	dot	cross	first	last	average
82	36.9%	80.1%	69.9%	75.0%	68.2%	35.1%	42.7%	58.2%
83	36.9%	80.1%	69.9%	75.0%	68.2%	35.1%	42.7%	58.2%
84	36.9%	80.1%	69.9%	74.8%	68.2%	35.1%	42.5%	58.2%
85	36.9%	80.1%	69.9%	74.8%	68.2%	35.1%	42.5%	58.2%
86	36.9%	80.1%	69.9%	74.8%	68.2%	35.3%	42.5%	58.2%
87	36.9%	80.1%	69.9%	74.8%	68.2%	35.3%	42.5%	58.2%
88	36.9%	80.1%	69.9%	74.8%	68.4%	35.3%	42.5%	58.2%
89	36.9%	80.1%	69.9%	74.8%	68.4%	35.3%	42.5%	58.2%
90	36.9%	80.1%	69.9%	74.8%	68.4%	35.3%	42.5%	58.2%
91	36.9%	80.1%	69.9%	74.8%	68.4%	35.3%	42.5%	58.2%
92	36.9%	80.1%	69.9%	74.8%	68.2%	35.3%	42.5%	58.2%
93	36.9%	80.1%	69.9%	74.8%	68.2%	35.3%	42.5%	58.2%
94	36.9%	80.1%	69.9%	74.8%	68.2%	35.3%	42.5%	58.2%
95	36.9%	80.1%	69.9%	74.8%	68.4%	35.3%	42.5%	58.2%
96	36.9%	80.1%	69.9%	74.8%	68.4%	35.3%	42.5%	58.2%
97	36.9%	80.1%	69.9%	74.8%	68.4%	35.3%	42.5%	58.2%
98	36.9%	80.1%	69.9%	74.8%	68.4%	35.3%	42.5%	58.2%
99	36.9%	80.1%	69.9%	74.8%	68.4%	35.3%	42.5%	58.2%
100	36.9%	80.1%	69.9%	74.8%	68.4%	35.3%	42.5%	58.2%

Table F-1: Percent correct of the cues: rank, without weighting, by threshold: partial data set

threshold	length	ascender	descender	dot	cross	first	last	average
1	33.4%	80.4%	71.5%	74.0%	66.1%	35.1%	40.2%	57.2%
2	34.4%	80.4%	71.2%	74.0%	65.6%	35.1%	40.2%	57.3%
3	37.6%	82.0%	70.3%	73.6%	67.7%	34.9%	41.2%	58.2%
4	38.9%	82.5%	70.8%	75.1%	67.5%	35.6%	41.8%	58.9%
5	37.7%	81.5%	70.5%	73.8%	67.9%	37.2%	41.7%	58.6%
6	37.7%	81.5%	70.5%	74.1%	67.4%	37.7%	41.8%	58.7%
7	38.2%	81.4%	70.8%	73.8%	68.0%	38.2%	42.0%	59.0%
8	36.9%	81.2%	71.7%	74.1%	68.4%	38.2%	41.7%	58.9%
9	37.7%	81.4%	71.3%	73.8%	67.9%	38.1%	41.2%	58.8%
10	38.6%	81.2%	71.7%	74.3%	68.5%	37.1%	41.5%	59.0%
11	38.1%	81.2%	71.2%	73.5%	69.2%	36.6%	41.8%	58.8%
12	38.1%	81.4%	71.0%	73.5%	69.4%	36.4%	41.7%	58.8%
13	37.9%	81.1%	70.8%	73.5%	68.4%	36.2%	41.5%	58.5%
14	38.4%	81.4%	71.0%	73.8%	69.2%	36.1%	41.4%	58.7%
15	37.9%	81.4%	71.3%	73.6%	69.0%	36.1%	41.5%	58.7%
16	38.4%	81.5%	71.7%	73.5%	69.0%	36.1%	42.0%	58.9%
17	38.4%	81.7%	71.7%	73.6%	68.9%	36.1%	41.5%	58.8%
18	38.4%	81.5%	71.2%	73.8%	69.0%	36.2%	41.5%	58.8%
19	38.2%	81.5%	71.2%	74.0%	68.5%	36.2%	41.2%	58.7%
20	38.2%	81.2%	71.0%	74.1%	68.7%	36.2%	41.2%	58.7%
21	37.7%	81.4%	71.3%	73.8%	68.5%	36.1%	41.2%	58.6%

threshold	length	ascender	descender	dot	cross	first	last	average
22	37.9%	81.2%	71.3%	73.6%	68.5%	35.9%	41.4%	58.6%
23	38.7%	81.4%	71.3%	73.8%	68.4%	35.9%	41.0%	58.6%
24	38.4%	81.4%	71.2%	73.6%	68.4%	36.1%	40.9%	58.6%
25	38.4%	81.2%	71.0%	73.6%	68.4%	36.1%	41.0%	58.5%
26	38.4%	81.1%	71.0%	73.6%	68.5%	36.1%	41.2%	58.6%
27	38.7%	81.2%	71.0%	73.6%	68.4%	36.4%	41.5%	58.7%
28	38.4%	81.2%	71.3%	73.6%	68.9%	36.2%	41.7%	58.8%
29	38.4%	81.2%	71.2%	73.6%	68.7%	36.2%	41.7%	58.7%
30	38.2%	81.2%	71.2%	73.6%	68.5%	36.4%	41.7%	58.7%
31	38.4%	81.2%	71.2%	73.6%	68.5%	36.4%	41.5%	58.7%
32	38.6%	81.2%	71.2%	74.0%	68.4%	36.2%	41.7%	58.7%
33	38.7%	81.2%	71.2%	74.0%	68.5%	36.4%	41.8%	58.8%
34	38.6%	81.1%	71.0%	73.8%	68.5%	36.4%	42.2%	58.8%
35	38.6%	81.2%	71.0%	73.8%	68.7%	36.6%	42.0%	58.8%
36	38.6%	81.2%	70.7%	73.8%	68.7%	36.6%	42.0%	58.8%
37	38.6%	81.2%	70.7%	74.1%	68.7%	36.4%	42.2%	58.8%
38	38.4%	81.2%	70.5%	74.0%	68.7%	36.4%	42.2%	58.8%
39	38.2%	81.4%	70.5%	74.0%	68.7%	36.2%	42.2%	58.7%
40	38.2%	81.4%	70.5%	74.0%	68.5%	36.2%	42.3%	58.7%
41	38.1%	81.4%	70.3%	74.0%	68.5%	36.4%	42.3%	58.7%
42	38.1%	81.4%	70.3%	74.0%	68.5%	36.2%	42.3%	58.7%
43	38.1%	81.4%	70.2%	74.0%	68.5%	36.2%	42.2%	58.6%
44	38.2%	81.2%	70.3%	74.0%	68.4%	36.2%	42.2%	58.6%
45	38.4%	81.2%	70.3%	74.1%	68.5%	36.4%	42.2%	58.7%
46	38.6%	81.2%	70.2%	74.1%	68.7%	36.1%	42.3%	58.7%
47	38.6%	81.2%	70.0%	74.1%	68.7%	36.1%	42.2%	58.7%
48	38.6%	81.1%	70.0%	74.1%	68.7%	36.1%	42.0%	58.6%
49	38.6%	81.1%	70.0%	74.1%	68.7%	36.1%	42.0%	58.6%
50	38.6%	81.1%	70.0%	74.1%	68.9%	36.1%	42.0%	58.7%
51	38.6%	81.2%	70.2%	74.0%	68.7%	36.4%	42.3%	58.8%
52	38.6%	81.1%	70.2%	74.0%	68.9%	36.4%	42.3%	58.8%
53	38.6%	81.2%	70.0%	74.0%	68.9%	36.4%	42.3%	58.8%
54	38.6%	81.2%	70.0%	74.0%	68.9%	36.2%	42.3%	58.7%

threshold	length	ascender	descender	dot	cross	first	last	average
55	38.6%	81.2%	70.0%	74.0%	68.9%	36.1%	42.2%	58.7%
56	38.6%	81.2%	70.0%	74.0%	68.7%	36.1%	42.2%	58.7%
57	38.6%	81.2%	69.9%	74.0%	68.9%	36.1%	42.3%	58.7%
58	38.6%	81.2%	69.9%	74.0%	68.9%	36.2%	42.3%	58.7%
59	38.7%	81.2%	69.9%	74.0%	68.9%	36.1%	42.3%	58.7%
60	38.7%	81.2%	69.9%	74.0%	68.9%	36.1%	42.3%	58.7%
61	38.7%	81.2%	69.9%	74.0%	68.9%	36.2%	42.3%	58.7%
62	38.7%	81.2%	69.9%	74.0%	68.7%	36.2%	42.3%	58.7%
63	38.6%	81.2%	69.9%	74.0%	68.7%	36.1%	42.5%	58.7%
64	38.4%	81.2%	69.9%	74.0%	68.7%	36.1%	42.5%	58.7%
65	38.4%	81.2%	69.9%	74.0%	68.7%	36.1%	42.5%	58.7%
66	38.6%	81.2%	69.9%	74.1%	68.7%	36.1%	42.5%	58.7%
67	38.6%	81.1%	69.9%	74.1%	68.7%	36.1%	42.5%	58.7%
68	38.6%	81.1%	69.9%	74.1%	68.7%	36.1%	42.5%	58.7%
69	38.6%	81.1%	69.9%	74.1%	68.7%	36.1%	42.5%	58.7%
70	38.6%	81.1%	69.9%	74.1%	68.7%	36.1%	42.5%	58.7%
71	38.6%	81.1%	69.9%	74.1%	68.7%	36.1%	42.5%	58.7%
72	38.6%	80.9%	69.9%	74.1%	68.7%	36.1%	42.5%	58.7%
73	38.6%	80.9%	69.9%	74.1%	68.7%	36.1%	42.7%	58.7%
74	38.6%	80.9%	69.9%	74.1%	68.7%	36.1%	42.7%	58.7%
75	38.6%	80.9%	69.9%	74.1%	68.7%	36.1%	42.7%	58.7%
76	38.6%	80.9%	69.9%	74.1%	68.7%	36.1%	42.7%	58.7%
77	38.6%	80.9%	69.9%	74.1%	68.7%	35.9%	42.5%	58.6%
78	38.6%	80.9%	69.9%	74.1%	68.7%	35.9%	42.7%	58.7%
79	38.6%	80.9%	69.9%	74.1%	68.7%	35.7%	42.7%	58.6%
80	38.6%	80.9%	69.9%	74.1%	68.7%	35.7%	42.7%	58.6%
81	38.6%	80.9%	69.9%	74.1%	68.7%	35.7%	42.7%	58.6%
82	38.6%	80.9%	69.9%	74.1%	68.7%	35.7%	42.7%	58.6%
83	38.6%	80.9%	69.9%	74.1%	68.7%	35.7%	42.7%	58.6%
84	38.6%	80.9%	69.9%	74.1%	68.7%	35.7%	42.7%	58.6%
85	38.6%	80.9%	69.9%	74.1%	68.7%	35.7%	42.7%	58.6%
86	38.6%	80.9%	69.9%	74.1%	68.7%	35.9%	42.7%	58.7%
87	38.6%	80.9%	69.9%	74.1%	68.7%	35.9%	42.7%	58.7%

threshold	length	ascender	descender	dot	cross	first	last	average
88	38.6%	80.9%	69.9%	74.1%	68.7%	35.9%	42.7%	58.7%
89	38.6%	80.9%	69.9%	74.1%	68.7%	35.9%	42.7%	58.7%
90	38.6%	80.9%	69.9%	74.1%	68.7%	35.9%	42.7%	58.7%
91	38.6%	80.9%	69.9%	74.1%	68.7%	35.9%	42.7%	58.7%
92	38.6%	80.9%	69.9%	74.1%	68.7%	35.9%	42.7%	58.7%
93	38.6%	80.9%	69.9%	74.1%	68.7%	35.9%	42.7%	58.7%
94	38.6%	80.9%	69.9%	74.1%	68.7%	35.9%	42.7%	58.7%
95	38.6%	80.9%	69.9%	74.1%	68.7%	35.9%	42.7%	58.7%
96	38.6%	80.9%	69.9%	74.1%	68.7%	35.9%	42.7%	58.7%
97	38.6%	80.9%	69.9%	74.1%	68.7%	35.9%	42.5%	58.6%
98	38.6%	80.9%	69.9%	74.1%	68.7%	35.9%	42.5%	58.6%
99	38.6%	80.9%	69.9%	74.1%	68.7%	35.9%	42.5%	58.6%
100	38.6%	80.9%	69.9%	74.1%	68.7%	35.9%	42.5%	58.6%

Table F-2: Percent correct of the cues: rank, with weighting, by threshold: partial data set

threshold	length	ascender	descender	dot	cross	first	last	average
100	33.4%	80.4%	71.5%	74.0%	66.1%	35.1%	40.2%	57.2%
99	33.6%	80.4%	71.5%	74.1%	66.2%	35.1%	40.2%	57.3%
98	33.6%	80.6%	71.5%	74.3%	66.1%	34.9%	40.2%	57.3%
97	33.8%	80.6%	71.5%	74.3%	66.1%	34.9%	40.2%	57.3%
96	33.9%	80.9%	71.3%	74.3%	66.1%	34.8%	40.5%	57.4%
95	34.8%	81.1%	71.3%	74.5%	66.1%	34.8%	40.9%	57.6%
94	34.6%	81.1%	71.3%	74.8%	66.4%	34.9%	40.9%	57.7%
93	34.9%	80.9%	71.3%	74.6%	65.9%	34.9%	41.2%	57.7%
92	35.1%	80.9%	71.3%	74.8%	65.9%	35.1%	41.4%	57.8%
91	35.3%	80.7%	71.3%	74.8%	66.1%	35.1%	41.4%	57.8%
90	35.4%	80.7%	71.3%	75.0%	66.1%	34.9%	41.5%	57.8%
89	35.4%	80.7%	71.2%	75.1%	66.1%	35.1%	41.5%	57.9%
88	35.1%	80.7%	71.0%	75.1%	65.7%	34.9%	41.4%	57.7%
87	35.3%	80.7%	70.8%	75.3%	65.9%	35.1%	41.5%	57.8%
86	34.9%	80.9%	70.8%	75.5%	66.1%	35.1%	41.4%	57.8%
85	35.3%	80.9%	70.8%	75.8%	66.4%	35.3%	41.4%	58.0%
84	35.7%	80.9%	70.8%	75.8%	66.4%	35.1%	41.4%	58.0%
83	35.4%	80.9%	70.8%	75.3%	66.4%	34.9%	41.4%	57.9%
82	35.7%	81.1%	70.8%	75.3%	66.2%	34.9%	41.4%	57.9%
81	35.9%	80.9%	70.7%	75.1%	66.1%	35.1%	40.9%	57.8%
80	36.4%	80.4%	70.8%	75.1%	66.1%	35.4%	40.9%	57.9%

threshold	length	ascender	descender	dot	cross	first	last	average
79	36.7%	80.6%	71.0%	75.3%	65.7%	35.6%	41.2%	58.0%
78	36.4%	80.6%	71.0%	75.3%	66.2%	35.7%	40.9%	58.0%
77	36.2%	80.7%	71.3%	75.0%	65.9%	35.7%	40.9%	58.0%
76	36.4%	80.6%	71.3%	74.3%	65.7%	35.4%	41.0%	57.8%
75	36.6%	80.6%	71.2%	74.5%	66.2%	35.6%	41.0%	57.9%
74	36.9%	80.6%	71.5%	74.0%	66.6%	35.3%	40.9%	57.9%
73	36.9%	80.9%	71.5%	75.1%	66.2%	35.4%	41.0%	58.2%
72	36.4%	81.1%	71.5%	75.1%	66.7%	34.9%	41.4%	58.2%
71	36.7%	81.1%	71.8%	74.5%	66.7%	35.1%	41.4%	58.2%
70	37.1%	81.1%	72.0%	75.0%	66.4%	34.9%	41.7%	58.3%
69	36.6%	81.1%	72.0%	75.0%	66.7%	34.9%	41.7%	58.3%
68	36.4%	80.9%	71.8%	74.3%	66.7%	34.8%	41.7%	58.1%
67	36.2%	81.1%	71.8%	74.1%	66.7%	34.8%	41.4%	58.0%
66	36.4%	80.7%	72.2%	73.8%	66.7%	34.6%	41.5%	58.0%
65	36.4%	80.6%	72.3%	74.5%	66.9%	34.8%	42.0%	58.2%
64	36.7%	80.2%	72.2%	74.3%	67.4%	34.8%	42.2%	58.2%
63	36.2%	80.4%	72.2%	74.1%	67.2%	35.1%	42.2%	58.2%
62	36.7%	80.2%	72.2%	74.0%	67.4%	34.8%	42.3%	58.2%
61	36.9%	80.2%	71.7%	73.5%	67.5%	35.3%	42.3%	58.2%
60	37.1%	79.9%	71.2%	73.3%	67.2%	35.1%	42.5%	58.0%
59	37.7%	80.4%	71.5%	74.0%	67.1%	35.6%	42.7%	58.4%
58	37.4%	80.2%	71.3%	73.8%	66.7%	35.9%	42.3%	58.2%
57	37.4%	80.7%	71.5%	74.0%	67.1%	36.4%	41.7%	58.4%
56	37.2%	81.1%	71.5%	74.5%	67.7%	36.1%	41.5%	58.5%
55	36.9%	81.1%	71.5%	74.5%	67.4%	36.1%	41.5%	58.4%
54	36.6%	80.9%	71.2%	75.5%	67.5%	35.7%	41.4%	58.4%
53	37.1%	80.6%	71.3%	75.1%	67.2%	35.7%	41.5%	58.4%
52	37.4%	80.9%	71.2%	75.0%	67.7%	35.6%	41.5%	58.5%
51	37.7%	80.9%	71.3%	75.3%	67.5%	36.1%	41.7%	58.6%
50	38.2%	80.9%	70.8%	75.6%	67.5%	36.4%	42.0%	58.8%
49	38.2%	80.6%	71.0%	75.9%	67.7%	37.4%	41.4%	58.9%
48	38.4%	80.6%	71.0%	75.3%	66.9%	37.1%	41.5%	58.7%
47	38.1%	80.9%	71.2%	74.6%	66.9%	36.9%	41.8%	58.6%

threshold	length	ascender	descender	dot	cross	first	last	average
46	38.4%	80.9%	71.0%	74.8%	67.1%	36.6%	41.8%	58.6%
45	38.4%	80.6%	70.7%	74.8%	67.1%	35.7%	41.5%	58.4%
44	38.2%	80.9%	70.7%	75.1%	67.7%	35.3%	41.7%	58.5%
43	37.9%	80.6%	70.3%	74.8%	67.9%	35.4%	41.5%	58.3%
42	38.2%	80.4%	70.0%	74.8%	68.2%	35.4%	41.8%	58.4%
41	37.9%	80.2%	70.3%	74.8%	69.5%	35.6%	41.8%	58.6%
40	38.2%	80.1%	70.5%	75.3%	69.4%	35.6%	42.5%	58.8%
39	38.1%	80.1%	70.5%	75.6%	68.7%	35.7%	42.5%	58.7%
38	38.2%	80.1%	70.5%	75.8%	68.9%	35.7%	42.5%	58.8%
37	37.7%	80.1%	70.2%	75.5%	69.0%	35.7%	42.5%	58.7%
36	37.4%	80.1%	70.3%	75.0%	68.7%	35.7%	42.7%	58.6%
35	37.4%	80.4%	70.2%	75.0%	69.5%	36.1%	42.5%	58.7%
34	37.9%	80.4%	70.0%	75.0%	69.4%	35.7%	42.8%	58.7%
33	37.2%	80.4%	70.3%	75.0%	69.0%	36.1%	43.0%	58.7%
32	37.1%	80.2%	70.3%	74.8%	68.7%	36.2%	43.2%	58.6%
31	36.7%	80.2%	70.7%	75.5%	69.0%	35.9%	42.8%	58.7%
30	37.2%	80.2%	70.7%	75.0%	69.0%	35.9%	43.2%	58.7%
29	37.4%	80.2%	70.8%	74.5%	68.7%	35.9%	42.7%	58.6%
28	37.7%	80.1%	70.8%	74.6%	68.5%	35.7%	42.8%	58.6%
27	37.2%	79.9%	70.8%	74.6%	68.0%	35.4%	42.3%	58.3%
26	37.6%	79.9%	70.5%	74.8%	68.4%	35.6%	42.2%	58.4%
25	37.9%	79.9%	70.0%	74.6%	68.7%	35.6%	42.3%	58.4%
24	37.9%	79.9%	69.7%	74.8%	68.2%	34.9%	42.2%	58.2%
23	37.2%	79.9%	70.0%	74.8%	67.9%	35.1%	42.0%	58.1%
22	36.9%	80.1%	69.9%	74.8%	68.4%	35.3%	42.5%	58.2%
21	36.9%	80.1%	69.9%	74.8%	68.4%	35.3%	42.5%	58.2%
20	36.9%	80.1%	69.9%	74.8%	68.4%	35.3%	42.5%	58.2%
19	36.9%	80.1%	69.9%	74.8%	68.4%	35.3%	42.5%	58.2%
18	36.9%	80.1%	69.9%	74.8%	68.4%	35.3%	42.5%	58.2%
17	36.9%	80.1%	69.9%	74.8%	68.4%	35.3%	42.5%	58.2%
16	36.9%	80.1%	69.9%	74.8%	68.4%	35.3%	42.5%	58.2%
15	36.9%	80.1%	69.9%	74.8%	68.4%	35.3%	42.5%	58.2%
14	36.9%	80.1%	69.9%	74.8%	68.4%	35.3%	42.5%	58.2%

threshold	length	ascender	descender	dot	cross	first	last	average
13	36.9%	80.1%	69.9%	74.8%	68.4%	35.3%	42.5%	58.2%
12	36.9%	80.1%	69.9%	74.8%	68.4%	35.3%	42.5%	58.2%
11	36.9%	80.1%	69.9%	74.8%	68.4%	35.3%	42.5%	58.2%
10	36.9%	80.1%	69.9%	74.8%	68.4%	35.3%	42.5%	58.2%
9	36.9%	80.1%	69.9%	74.8%	68.4%	35.3%	42.5%	58.2%
8	36.9%	80.1%	69.9%	74.8%	68.4%	35.3%	42.5%	58.2%
7	36.9%	80.1%	69.9%	74.8%	68.4%	35.3%	42.5%	58.2%
6	36.9%	80.1%	69.9%	74.8%	68.4%	35.3%	42.5%	58.2%
5	36.9%	80.1%	69.9%	74.8%	68.4%	35.3%	42.5%	58.2%
4	36.9%	80.1%	69.9%	74.8%	68.4%	35.3%	42.5%	58.2%
3	36.9%	80.1%	69.9%	74.8%	68.4%	35.3%	42.5%	58.2%
2	36.9%	80.1%	69.9%	74.8%	68.4%	35.3%	42.5%	58.2%

Table F-3: Percent correct of the cues: score, without weighting, by threshold: partial data set

threshold	length	ascender	descender	dot	cross	first	last	average
100	33.4%	80.4%	71.5%	74.0%	66.1%	35.1%	40.2%	57.2%
99	33.4%	80.4%	71.5%	74.0%	66.1%	35.1%	40.2%	57.2%
98	33.6%	80.6%	71.5%	74.0%	66.1%	34.9%	40.2%	57.3%
97	33.8%	80.6%	71.5%	74.0%	66.1%	34.9%	40.2%	57.3%
96	33.9%	80.9%	71.3%	73.8%	66.1%	34.8%	40.5%	57.3%
95	34.8%	80.7%	71.3%	73.8%	66.1%	34.8%	40.7%	57.4%
94	34.6%	80.7%	71.3%	74.1%	66.1%	34.9%	40.9%	57.5%
93	34.9%	80.7%	71.3%	74.1%	66.1%	34.9%	41.2%	57.6%
92	35.1%	80.9%	71.3%	74.1%	66.2%	34.9%	41.4%	57.7%
91	35.3%	80.7%	71.3%	74.3%	66.4%	34.9%	41.4%	57.8%
90	35.4%	80.7%	71.3%	74.5%	66.4%	34.9%	41.5%	57.8%
89	35.3%	80.7%	71.2%	74.5%	66.4%	35.1%	41.5%	57.8%
88	35.1%	80.7%	71.0%	74.5%	66.1%	34.9%	41.4%	57.7%
87	35.1%	80.7%	70.8%	74.5%	66.1%	34.9%	41.5%	57.7%
86	34.9%	80.9%	70.8%	74.5%	66.2%	35.1%	41.4%	57.7%
85	35.3%	80.9%	70.8%	74.6%	66.4%	35.1%	41.4%	57.8%
84	35.7%	80.7%	70.8%	74.6%	66.7%	35.1%	41.4%	57.9%
83	35.3%	80.9%	70.8%	74.3%	66.7%	34.9%	41.2%	57.7%
82	35.3%	81.1%	71.0%	74.5%	66.6%	34.9%	41.2%	57.8%
81	35.1%	80.9%	70.8%	74.6%	66.4%	34.9%	40.7%	57.6%
80	35.7%	80.4%	70.8%	74.8%	66.4%	35.1%	40.5%	57.7%
79	35.4%	80.6%	71.0%	75.0%	66.1%	35.3%	40.7%	57.7%
78	35.7%	80.6%	71.0%	75.0%	66.4%	35.4%	40.4%	57.8%

threshold	length	ascender	descender	dot	cross	first	last	average
77	36.4%	80.4%	71.3%	74.6%	66.1%	35.7%	40.4%	57.8%
76	36.2%	80.4%	71.3%	74.3%	66.1%	35.4%	40.5%	57.8%
75	36.9%	80.4%	71.2%	74.3%	66.6%	35.7%	40.5%	57.9%
74	37.1%	80.6%	71.3%	73.3%	66.6%	35.3%	40.9%	57.8%
73	37.2%	80.9%	71.2%	74.3%	66.4%	35.4%	40.9%	58.0%
72	36.9%	81.1%	71.2%	74.5%	66.7%	35.1%	41.2%	58.1%
71	36.7%	80.7%	71.5%	74.1%	66.7%	35.1%	41.2%	58.0%
70	37.1%	80.9%	71.5%	74.5%	66.4%	34.9%	41.5%	58.1%
69	36.9%	80.9%	71.5%	74.3%	66.7%	34.9%	41.5%	58.1%
68	37.1%	80.9%	71.3%	73.8%	66.7%	34.9%	41.5%	58.0%
67	36.9%	81.1%	71.8%	73.8%	66.7%	34.9%	41.4%	58.1%
66	37.1%	80.7%	72.2%	73.5%	66.9%	34.8%	41.4%	58.1%
65	36.7%	80.4%	72.2%	73.8%	66.9%	34.8%	41.5%	58.0%
64	36.6%	80.2%	72.3%	73.5%	67.4%	34.8%	41.8%	58.1%
63	36.4%	80.6%	71.8%	73.3%	67.2%	35.1%	42.2%	58.1%
62	36.9%	80.6%	71.8%	73.1%	67.2%	34.8%	42.0%	58.1%
61	36.9%	80.2%	71.5%	73.1%	67.2%	35.3%	42.0%	58.0%
60	36.9%	80.1%	71.2%	73.3%	67.2%	35.3%	42.3%	58.0%
59	37.6%	80.4%	71.5%	73.6%	67.2%	35.6%	42.3%	58.3%
58	37.2%	80.2%	71.5%	73.8%	66.7%	35.9%	42.2%	58.2%
57	37.4%	81.1%	71.7%	73.8%	66.7%	36.4%	41.4%	58.3%
56	37.2%	81.1%	71.8%	74.0%	67.4%	36.1%	41.2%	58.4%
55	36.6%	81.1%	71.7%	74.5%	67.1%	36.1%	41.7%	58.4%
54	37.4%	80.9%	71.3%	74.3%	67.9%	36.1%	41.5%	58.5%
53	37.4%	80.7%	71.3%	74.8%	67.1%	35.7%	41.5%	58.4%
52	37.7%	81.1%	71.2%	74.5%	67.2%	35.6%	41.5%	58.4%
51	37.9%	80.9%	71.3%	74.6%	67.5%	35.9%	41.5%	58.5%
50	38.2%	81.2%	70.7%	75.0%	67.4%	36.2%	41.7%	58.6%
49	37.9%	80.9%	70.8%	74.6%	67.7%	37.2%	41.2%	58.6%
48	37.7%	81.4%	71.2%	74.0%	67.2%	37.1%	41.4%	58.6%
47	38.1%	81.7%	71.2%	73.5%	67.5%	37.4%	41.5%	58.7%
46	38.2%	81.5%	70.8%	73.8%	67.1%	37.4%	41.5%	58.6%
45	38.4%	81.2%	70.8%	73.5%	67.4%	36.7%	42.0%	58.6%

threshold	length	ascender	descender	dot	cross	first	last	average
44	38.4%	81.1%	70.7%	73.6%	68.2%	36.6%	41.7%	58.6%
43	38.2%	81.1%	70.7%	73.8%	68.2%	36.6%	41.4%	58.6%
42	38.7%	81.1%	70.5%	74.1%	67.7%	36.6%	41.5%	58.6%
41	38.6%	80.2%	70.7%	73.8%	68.9%	36.6%	41.8%	58.6%
40	38.9%	80.6%	70.3%	74.1%	68.4%	36.4%	42.8%	58.8%
39	38.6%	80.6%	70.7%	74.3%	68.4%	36.4%	43.0%	58.8%
38	38.7%	80.4%	70.8%	74.5%	68.9%	36.2%	43.2%	59.0%
37	38.7%	80.6%	70.2%	74.3%	69.2%	36.4%	43.2%	58.9%
36	38.4%	80.6%	70.3%	74.3%	68.9%	36.4%	43.0%	58.8%
35	38.9%	80.9%	70.0%	74.5%	69.4%	36.6%	43.0%	59.0%
34	38.7%	80.9%	70.0%	74.5%	69.4%	36.2%	43.0%	59.0%
33	38.4%	80.9%	70.3%	74.1%	69.4%	36.4%	42.8%	58.9%
32	38.6%	80.9%	70.3%	73.6%	69.2%	36.6%	42.7%	58.8%
31	38.6%	81.1%	70.2%	74.5%	69.2%	36.9%	42.8%	59.0%
30	38.7%	81.1%	70.2%	74.1%	69.0%	36.6%	42.8%	58.9%
29	38.6%	81.1%	70.3%	73.8%	69.0%	36.7%	42.7%	58.9%
28	38.9%	80.9%	70.5%	73.8%	69.0%	36.6%	42.5%	58.9%
27	38.6%	80.7%	70.5%	74.0%	68.7%	36.1%	42.3%	58.7%
26	39.0%	80.7%	70.3%	74.1%	68.9%	36.2%	42.0%	58.8%
25	39.4%	80.7%	70.0%	74.6%	68.7%	36.2%	42.2%	58.8%
24	39.2%	80.7%	69.9%	74.5%	68.7%	35.7%	42.3%	58.7%
23	38.7%	80.7%	69.7%	74.5%	68.7%	35.9%	42.3%	58.6%
22	38.6%	80.9%	69.9%	74.1%	68.7%	35.9%	42.5%	58.6%
21	38.6%	80.9%	69.9%	74.1%	68.7%	35.9%	42.5%	58.6%
20	38.6%	80.9%	69.9%	74.1%	68.7%	35.9%	42.5%	58.6%
19	38.6%	80.9%	69.9%	74.1%	68.7%	35.9%	42.5%	58.6%
18	38.6%	80.9%	69.9%	74.1%	68.7%	35.9%	42.5%	58.6%
17	38.6%	80.9%	69.9%	74.1%	68.7%	35.9%	42.5%	58.6%
16	38.6%	80.9%	69.9%	74.1%	68.7%	35.9%	42.5%	58.6%
15	38.6%	80.9%	69.9%	74.1%	68.7%	35.9%	42.5%	58.6%
14	38.6%	80.9%	69.9%	74.1%	68.7%	35.9%	42.5%	58.6%
13	38.6%	80.9%	69.9%	74.1%	68.7%	35.9%	42.5%	58.6%
12	38.6%	80.9%	69.9%	74.1%	68.7%	35.9%	42.5%	58.6%

threshold	length	ascender	descender	dot	cross	first	last	average
11	38.6%	80.9%	69.9%	74.1%	68.7%	35.9%	42.5%	58.6%
10	38.6%	80.9%	69.9%	74.1%	68.7%	35.9%	42.5%	58.6%
9	38.6%	80.9%	69.9%	74.1%	68.7%	35.9%	42.5%	58.6%
8	38.6%	80.9%	69.9%	74.1%	68.7%	35.9%	42.5%	58.6%
7	38.6%	80.9%	69.9%	74.1%	68.7%	35.9%	42.5%	58.6%
6	38.6%	80.9%	69.9%	74.1%	68.7%	35.9%	42.5%	58.6%
5	38.6%	80.9%	69.9%	74.1%	68.7%	35.9%	42.5%	58.6%
4	38.6%	80.9%	69.9%	74.1%	68.7%	35.9%	42.5%	58.6%
3	38.6%	80.9%	69.9%	74.1%	68.7%	35.9%	42.5%	58.6%
2	38.6%	80.9%	69.9%	74.1%	68.7%	35.9%	42.5%	58.6%

Table F-4: Percent correct of the cues: score, with weighting, by threshold: partial data set

In Table F-5 to Table F-8 results are shown only for the chosen threshold used to limit the number of candidates. Results are given by the combination of methods used to calculate the results. A dash means that the combination of methods cannot be used for that cue.

Mean, median and mode were the three ways used to calculate values (see section 4.3.2).

Average, reduce and initial were the three ways used to force a single-valued outcome (see section 4.3.3).

method	length	ascender	descender	dot	cross	first	last
mean	37.2%	80.2%	69.2%	75.6%	67.2%	-	-
median, average	36.4%	80.2%	69.2%	75.6%	67.2%	-	-
median, reduce	36.6%	80.7%	70.5%	73.1%	67.4%	-	-
median, initial	34.8%	80.9%	72.0%	73.6%	68.0%	-	-
mode, average	36.2%	80.2%	69.2%	75.6%	67.2%	-	-
mode, reduce	35.4%	80.7%	70.5%	73.1%	67.4%	37.6%	42.0%
mode, initial	35.9%	80.9%	72.0%	73.6%	68.0%	38.4%	41.0%

Table F-5: Percent correct of the cues: rank, without weighting, by combination of methods: partial data set

method	length	ascender	descender	dot	cross	first	last
mean	37.1%	81.4%	70.3%	73.8%	68.0%	-	-
median, average	38.2%	81.4%	70.3%	73.8%	68.0%	-	-
median, reduce	38.2%	81.4%	70.7%	73.8%	67.9%	-	-
median, initial	38.2%	81.4%	70.8%	73.8%	67.9%	-	-
mode, average	37.1%	81.4%	70.3%	73.8%	68.0%	-	-
mode, reduce	36.9%	81.4%	70.7%	73.8%	67.9%	38.2%	42.0%
mode, initial	36.9%	81.4%	70.8%	73.8%	67.9%	38.2%	42.0%

Table F-6: Percent correct of the cues: rank, with weighting, by combination of methods: partial data set

method	length	ascender	descender	dot	cross	first	last
mean	37.1%	80.6%	68.9%	75.9%	67.2%	-	-
median, average	37.2%	80.6%	68.9%	75.9%	67.2%	-	-
median, reduce	38.2%	80.2%	70.5%	74.6%	67.7%	-	-
median, initial	37.7%	80.4%	71.0%	74.5%	67.2%	-	-
mode, average	38.1%	80.6%	68.9%	75.9%	67.2%	-	-
mode, reduce	37.6%	80.2%	70.5%	74.6%	67.7%	36.7%	41.4%
mode, initial	37.1%	80.4%	71.0%	74.5%	67.2%	37.4%	40.9%

Table F-7: Percent correct of the cues: score, without weighting, combination of methods: partial data set

method	length	ascender	descender	dot	cross	first	last
mean	38.4%	80.9%	69.4%	74.5%	69.4%	-	-
median, average	38.9%	80.9%	69.4%	74.5%	69.4%	-	-
median, reduce	38.9%	80.9%	70.0%	74.5%	69.2%	-	-
median, initial	38.9%	80.9%	70.0%	74.5%	69.2%	-	-
mode, average	36.4%	80.9%	69.4%	74.5%	69.4%	-	-
mode, reduce	36.4%	80.9%	70.0%	74.5%	69.2%	36.6%	43.0%
mode, initial	36.4%	80.9%	70.0%	74.5%	69.2%	36.6%	43.0%

Table F-8: Percent correct of the cues: score, with weighting, by combination of methods: partial data set

Appendix G: Merging the Word Level Method with the Pattern Recognizer: Initial Results

The list of words generated by the word level method was merged with the list of words generated by the pattern recognizer (see section 4.4.2). Only higher ranked candidates from the list of alternatives generated by the pattern recognizer were used. In Table G-1 results are given for a range of thresholds used to limit the number of alternatives.

confi-dence score	rank 1	rank 2	rank 3	rank 4	rank 5	rank 6	rank 7	rank 8	rank 9	rank 10	rank 100
100	61.5%	64.7%	66.3%	67.0%	69.4%	70.1%	70.8%	71.6%	72.1%	73.1%	97.1%
99	61.5%	65.0%	66.8%	67.5%	69.8%	70.5%	71.2%	71.8%	72.8%	73.7%	97.1%
98	61.5%	65.3%	67.3%	67.8%	70.2%	70.9%	71.5%	72.5%	73.5%	74.7%	97.1%
97	61.5%	65.6%	67.6%	68.2%	70.6%	71.2%	72.3%	73.2%	74.5%	75.1%	97.1%
96	61.5%	65.7%	68.0%	68.6%	71.0%	71.9%	72.9%	74.2%	75.0%	75.6%	97.1%
95	61.5%	66.0%	68.3%	68.9%	71.7%	72.4%	73.8%	74.6%	75.4%	76.1%	97.1%
94	61.6%	66.4%	68.6%	69.6%	72.1%	73.4%	74.3%	75.1%	75.9%	76.5%	97.1%
93	61.7%	66.8%	69.1%	70.0%	73.0%	73.9%	74.8%	75.5%	76.3%	76.8%	97.1%
92	62.0%	67.2%	69.4%	71.1%	73.5%	74.4%	75.3%	76.0%	76.6%	77.1%	97.1%
91	62.4%	67.6%	70.6%	71.6%	74.0%	75.0%	75.7%	76.3%	76.9%	77.7%	97.1%
90	62.5%	68.5%	71.1%	72.0%	74.5%	75.4%	76.1%	76.5%	77.4%	78.1%	97.1%
89	63.2%	69.0%	71.5%	72.4%	74.8%	75.7%	76.3%	77.0%	77.9%	78.6%	97.1%
88	63.5%	69.5%	71.9%	72.9%	75.1%	75.9%	76.8%	77.5%	78.3%	78.8%	97.1%
87	64.0%	69.9%	72.4%	73.3%	75.4%	76.4%	77.1%	78.0%	78.6%	79.3%	97.1%
86	64.1%	70.2%	72.8%	73.6%	75.8%	76.6%	77.7%	78.3%	79.1%	79.7%	97.1%
85	64.4%	70.6%	73.1%	74.0%	76.0%	77.2%	78.0%	78.8%	79.4%	79.9%	97.1%
84	64.7%	71.0%	73.5%	74.3%	76.5%	77.5%	78.5%	79.1%	79.7%	80.5%	97.1%
83	64.9%	71.3%	73.7%	74.8%	76.9%	78.0%	78.7%	79.4%	80.2%	80.8%	97.1%

confi-dence score	rank 1	rank 2	rank 3	rank 4	rank 5	rank 6	rank 7	rank 8	rank 9	rank 10	rank 100
82	65.2%	71.5%	74.2%	75.1%	77.4%	78.3%	79.0%	79.9%	80.6%	81.3%	97.1%
81	65.3%	71.9%	74.4%	75.6%	77.6%	78.5%	79.5%	80.2%	81.0%	81.7%	97.1%
80	65.7%	72.1%	74.9%	75.7%	78.0%	79.0%	79.8%	80.7%	81.5%	82.1%	97.1%
79	66.0%	72.5%	75.0%	76.0%	78.5%	79.2%	80.3%	81.1%	81.9%	82.3%	97.1%
78	66.4%	72.7%	75.3%	76.4%	78.8%	79.8%	80.8%	81.6%	82.1%	82.4%	97.1%
77	66.5%	72.9%	75.6%	76.9%	79.3%	80.2%	81.2%	81.8%	82.2%	82.6%	97.1%
76	66.7%	73.3%	76.0%	77.5%	79.7%	80.5%	81.4%	81.9%	82.4%	82.8%	97.1%
75	67.0%	73.5%	76.7%	77.9%	80.1%	80.8%	81.5%	82.0%	82.6%	83.2%	97.1%
74	67.3%	74.3%	77.0%	78.2%	80.3%	80.8%	81.6%	82.2%	83.0%	83.6%	97.1%
73	67.9%	74.6%	77.4%	78.4%	80.5%	80.9%	81.8%	82.6%	83.3%	83.8%	97.1%
72	68.2%	75.0%	77.6%	78.5%	80.6%	81.2%	82.2%	82.9%	83.5%	84.0%	97.1%
71	68.5%	75.1%	77.7%	78.6%	80.8%	81.5%	82.4%	83.1%	83.7%	84.2%	97.1%
70	68.5%	75.3%	77.9%	78.9%	81.1%	81.7%	82.6%	83.4%	83.9%	84.4%	97.1%
69	68.7%	75.5%	78.2%	79.3%	81.4%	82.0%	82.8%	83.5%	84.1%	84.6%	97.1%
68	68.8%	75.7%	78.5%	79.6%	81.5%	82.2%	83.0%	83.8%	84.4%	84.9%	97.1%
67	69.0%	76.1%	78.8%	79.7%	81.7%	82.3%	83.1%	84.0%	84.6%	85.0%	97.1%
66	69.1%	76.3%	78.9%	79.9%	81.8%	82.5%	83.4%	84.2%	84.8%	85.1%	97.1%
65	69.3%	76.5%	79.1%	80.0%	82.0%	82.7%	83.6%	84.4%	84.9%	85.3%	97.1%
64	69.5%	76.6%	79.2%	80.2%	82.2%	83.0%	83.8%	84.5%	85.1%	85.4%	97.1%
63	69.7%	76.7%	79.4%	80.5%	82.5%	83.1%	83.9%	84.7%	85.2%	85.5%	97.1%
62	69.7%	76.9%	79.6%	80.8%	82.7%	83.2%	84.1%	84.8%	85.3%	85.7%	97.1%
61	69.9%	77.1%	79.8%	81.0%	82.8%	83.4%	84.2%	85.0%	85.4%	85.7%	97.1%
60	70.0%	77.3%	80.0%	81.0%	82.9%	83.5%	84.3%	85.1%	85.5%	85.9%	97.1%
59	70.2%	77.4%	80.1%	81.1%	83.0%	83.7%	84.4%	85.1%	85.6%	85.9%	97.1%
58	70.3%	77.5%	80.2%	81.2%	83.1%	83.8%	84.5%	85.3%	85.7%	86.0%	97.1%
57	70.4%	77.6%	80.3%	81.4%	83.2%	83.9%	84.5%	85.3%	85.7%	86.1%	97.1%
56	70.6%	77.7%	80.4%	81.5%	83.3%	83.9%	84.6%	85.4%	85.8%	86.2%	97.1%
55	70.6%	77.8%	80.6%	81.6%	83.4%	84.0%	84.7%	85.5%	85.9%	86.2%	97.1%
54	70.6%	77.9%	80.6%	81.7%	83.5%	84.1%	84.8%	85.6%	85.8%	86.1%	97.1%
53	70.7%	77.9%	80.7%	81.7%	83.5%	84.2%	84.9%	85.5%	85.8%	86.2%	97.1%
52	70.8%	77.9%	80.6%	81.7%	83.6%	84.2%	84.9%	85.5%	85.9%	86.3%	97.1%
51	70.8%	78.0%	80.6%	81.9%	83.6%	84.2%	84.9%	85.6%	86.0%	86.4%	97.1%
50	70.8%	78.0%	80.8%	81.9%	83.6%	84.2%	85.0%	85.7%	86.1%	86.4%	97.1%
49	70.9%	78.1%	80.8%	81.9%	83.6%	84.3%	85.0%	85.8%	86.1%	86.5%	97.1%
48	70.9%	78.0%	80.8%	81.9%	83.7%	84.3%	85.1%	85.8%	86.2%	86.6%	97.1%
47	70.9%	78.0%	80.8%	82.0%	83.8%	84.4%	85.1%	85.8%	86.3%	86.6%	97.1%
46	70.9%	78.0%	80.8%	82.0%	83.9%	84.4%	85.2%	85.9%	86.3%	86.7%	97.1%
45	70.9%	78.1%	80.8%	82.1%	83.9%	84.4%	85.3%	86.0%	86.3%	86.7%	97.1%
44	70.9%	78.1%	80.9%	82.2%	84.0%	84.5%	85.3%	86.0%	86.3%	86.7%	97.1%

confi-dence score	rank 1	rank 2	rank 3	rank 4	rank 5	rank 6	rank 7	rank 8	rank 9	rank 10	rank 100
43	70.9%	78.2%	80.9%	82.1%	84.0%	84.6%	85.4%	85.9%	86.3%	86.7%	97.1%
42	71.0%	78.2%	80.9%	82.2%	84.1%	84.6%	85.3%	85.9%	86.3%	86.7%	97.1%
41	71.0%	78.2%	81.0%	82.2%	84.1%	84.6%	85.3%	85.9%	86.3%	86.7%	97.1%
40	71.0%	78.2%	81.0%	82.3%	84.0%	84.6%	85.3%	85.9%	86.3%	86.7%	97.1%
39	71.0%	78.2%	81.1%	82.3%	84.0%	84.6%	85.3%	86.0%	86.3%	86.7%	97.1%
38	71.0%	78.3%	81.0%	82.2%	84.0%	84.6%	85.3%	86.0%	86.3%	86.7%	97.1%
37	71.0%	78.3%	81.0%	82.1%	84.0%	84.5%	85.3%	86.0%	86.3%	86.7%	97.1%
36	71.1%	78.3%	81.0%	82.1%	84.0%	84.5%	85.3%	86.0%	86.4%	86.7%	97.1%
35	71.0%	78.3%	81.0%	82.1%	84.0%	84.6%	85.3%	86.0%	86.3%	86.7%	97.1%
34	71.0%	78.3%	81.0%	82.1%	84.1%	84.6%	85.3%	86.0%	86.3%	86.7%	97.1%
33	71.0%	78.3%	80.9%	82.1%	84.1%	84.6%	85.3%	86.0%	86.3%	86.7%	97.1%
32	71.0%	78.3%	80.9%	82.1%	84.1%	84.6%	85.3%	86.0%	86.3%	86.7%	97.1%
31	70.9%	78.3%	80.9%	82.1%	84.1%	84.6%	85.3%	86.0%	86.3%	86.7%	97.1%
30	70.9%	78.3%	80.9%	82.1%	84.1%	84.6%	85.3%	86.0%	86.3%	86.7%	97.1%
29	70.9%	78.3%	81.0%	82.1%	84.1%	84.6%	85.4%	86.0%	86.3%	86.7%	97.1%
28	70.8%	78.3%	81.0%	82.1%	84.1%	84.7%	85.4%	86.0%	86.3%	86.7%	97.1%
27	70.8%	78.3%	80.9%	82.1%	84.1%	84.7%	85.4%	86.0%	86.3%	86.7%	97.1%
26	70.8%	78.2%	80.9%	82.1%	84.1%	84.7%	85.4%	86.0%	86.3%	86.7%	97.1%
25	70.8%	78.1%	80.9%	82.1%	84.1%	84.7%	85.4%	86.0%	86.3%	86.7%	97.1%
24	70.8%	78.1%	80.9%	82.1%	84.1%	84.7%	85.4%	86.0%	86.3%	86.7%	97.1%
23	70.8%	78.1%	80.9%	82.1%	84.1%	84.7%	85.4%	86.0%	86.3%	86.7%	97.1%
22	70.8%	78.1%	80.9%	82.1%	84.1%	84.7%	85.4%	86.0%	86.3%	86.7%	97.1%
21	70.8%	78.1%	80.9%	82.1%	84.1%	84.7%	85.4%	86.0%	86.3%	86.7%	97.1%
20	70.8%	78.1%	80.9%	82.1%	84.1%	84.7%	85.4%	86.0%	86.3%	86.7%	97.1%
19	70.8%	78.1%	80.9%	82.1%	84.1%	84.7%	85.4%	86.0%	86.3%	86.7%	97.1%
18	70.8%	78.1%	80.9%	82.1%	84.1%	84.7%	85.4%	86.0%	86.3%	86.7%	97.1%
17	70.8%	78.1%	80.9%	82.1%	84.1%	84.7%	85.4%	86.0%	86.3%	86.7%	97.1%
16	70.8%	78.1%	80.9%	82.1%	84.1%	84.7%	85.4%	86.0%	86.3%	86.7%	97.1%
15	70.8%	78.1%	80.9%	82.1%	84.1%	84.7%	85.4%	86.0%	86.3%	86.7%	97.1%
14	70.8%	78.1%	80.9%	82.1%	84.1%	84.7%	85.4%	86.0%	86.3%	86.7%	97.1%
13	70.8%	78.1%	80.9%	82.1%	84.1%	84.7%	85.4%	86.0%	86.3%	86.7%	97.1%
12	70.8%	78.1%	80.9%	82.1%	84.1%	84.7%	85.4%	86.0%	86.3%	86.7%	97.1%
11	70.8%	78.1%	80.9%	82.1%	84.1%	84.7%	85.4%	86.0%	86.3%	86.7%	97.1%
10	70.8%	78.1%	80.9%	82.1%	84.1%	84.7%	85.4%	86.0%	86.3%	86.7%	97.1%
9	70.8%	78.1%	80.9%	82.1%	84.1%	84.7%	85.4%	86.0%	86.3%	86.7%	97.1%
8	70.8%	78.1%	80.9%	82.1%	84.1%	84.7%	85.4%	86.0%	86.3%	86.7%	97.1%
7	70.8%	78.1%	80.9%	82.1%	84.1%	84.7%	85.4%	86.0%	86.3%	86.7%	97.1%
6	70.8%	78.1%	80.9%	82.1%	84.1%	84.7%	85.4%	86.0%	86.3%	86.7%	97.1%
5	70.8%	78.1%	80.9%	82.1%	84.1%	84.7%	85.4%	86.0%	86.3%	86.7%	97.1%

confi-dence score	rank 1	rank 2	rank 3	rank 4	rank 5	rank 6	rank 7	rank 8	rank 9	rank 10	rank 100
4	70.8%	78.1%	80.9%	82.1%	84.1%	84.7%	85.4%	86.0%	86.3%	86.7%	97.1%
3	70.8%	78.1%	80.9%	82.1%	84.1%	84.7%	85.4%	86.0%	86.3%	86.7%	97.1%
2	70.8%	78.1%	80.9%	82.1%	84.1%	84.7%	85.4%	86.0%	86.3%	86.7%	97.1%
1	70.8%	78.1%	80.9%	82.1%	84.1%	84.7%	85.4%	86.0%	86.3%	86.7%	97.1%
0	70.8%	78.1%	80.9%	82.1%	84.1%	84.7%	85.4%	86.0%	86.3%	86.7%	97.1%

Table G-1: Merging the word level method with the pattern recognizer. Percent target word recognized at, or above, rank, with results given by threshold: complete 200 word data set, excluding catastrophic failures.

A dummy confidence score was given to the words in the candidate list generated by the word level method. Table G-2 shows the effect of a range of values for this dummy confidence score.

thresh-old	rank 1	rank 2	rank 3	rank 4	rank 5	rank 6	rank 7	rank 8	rank 9	rank 10	rank 100
1	71.1%	75.4%	77.1%	77.8%	79.5%	80.4%	81.2%	81.8%	82.2%	82.6%	94.5
2	71.1%	78.2%	80.9%	81.2%	83.2%	83.5%	84.4%	84.9%	85.2%	85.6%	96.4
3	71.1%	78.3%	81.0%	82.1%	84.0%	84.5%	85.3%	86.0%	86.5%	86.9%	97.1
4	71.1%	78.3%	80.9%	81.3%	84.0%	84.4%	85.4%	85.9%	86.6%	86.9%	97.1
5	71.1%	78.3%	80.9%	81.3%	83.4%	84.4%	85.2%	86.2%	86.7%	86.9%	97.1
6	71.1%	78.3%	80.9%	81.3%	83.4%	83.7%	85.0%	85.8%	86.5%	86.8%	97.0
7	71.1%	78.3%	80.9%	81.3%	83.4%	83.7%	84.5%	85.7%	86.5%	86.8%	97.0
8	71.1%	78.3%	80.9%	81.3%	83.4%	83.7%	84.5%	85.2%	86.4%	86.7%	96.9
9	71.1%	78.3%	80.9%	81.3%	83.4%	83.7%	84.5%	85.2%	85.9%	86.8%	96.9
10	71.1%	78.3%	80.9%	81.3%	83.4%	83.7%	84.5%	85.2%	85.9%	86.3%	96.9

Table G-2: Merging the word level method with the pattern recognizer. Percent target word recognized at, or above, rank, with results given by threshold: complete 200 word data set, excluding catastrophic failures.

Appendix H: Cue Derivation- Letter Associations

Table H-1 and Table H-2 show the confusion matrices for letter associations (see section 5.2.5). Each matrix has been split into two because of size considerations. The confusion matrices were constructed by comparing the candidates from the complete data set with their target values. The confusion matrices show the percentage chance of a confusion occurring. The matrices are shown from confusion to target. This is the way that they are used in practice. The matrices have been left blank where no confusion exists. A dash has been placed in the matrices whenever a letter coincides with itself.

	a	b	c	d	e	f	g	h	i	j	k	l	m
a	-	1%	8%	6%	7%	1%	1%	7%	6%		1%	1%	9%
b	9%	-	4%	1%	1%	14%	1%	27%	1%		1%	3%	1%
c	27%	1%	-	6%	11%	4%	1%	4%	3%		1%	1%	1%
d	27%	1%	9%	-	30%		2%	5%	1%			1%	1%
e	31%	1%	7%	4%	-	5%	3%	3%	3%				
f	8%	1%	3%	3%	3%	-	3%	12%	2%	2%	1%	1%	5%
g	17%		1%	2%	1%		-	1%	1%				
h	6%	1%	2%	1%	1%	3%	1%	-	3%			1%	5%
i	16%	1%	6%	2%	1%	5%	1%	8%	-	1%	1%	1%	9%
j	6%		3%	1%	1%	17%	2%	2%	7%	-		1%	7%
k	7%		10%	3%	11%	4%		33%			-	1%	
l	9%	1%	2%	3%	4%	11%	1%	13%	2%	1%	1%	-	3%
m	13%	1%	3%	2%	2%	1%	1%	9%	10%	1%		1%	-
n	13%	1%	3%	1%	3%	1%		15%	3%				11%
o	28%	1%	8%	3%	2%	2%	1%	3%	3%			1%	8%
p	1%	9%	1%	1%	1%	12%	1%	6%	6%	2%			17%
q	59%				3%		21%		3%				
r	8%	1%	3%	2%	2%	5%	2%	6%	3%	1%		1%	7%
s	10%	4%	2%	1%	1%	9%	2%	2%	6%	1%			5%
t	9%	1%	5%	3%	3%	19%	1%	18%	2%	1%		1%	2%
u	19%	1%	1%	1%	9%	1%		3%	8%			1%	13%
v	10%	1%	4%	1%	2%	2%	1%	3%	4%	1%		1%	10%
w	13%		1%	1%	1%	1%	1%	10%	3%	1%		1%	8%
x													
y		8%	8%			8%	23%	8%	8%				8%
z								7%					

	n	o	p	q	r	s	t	u	v	w	x	y	z
a	13%	5%	1%	2%	1%	4%	4%	10%	2%	13%			1%
b	4%	5%	2%	1%	1%	1%	10%	6%	3%	6%		2%	1%
c	4%	3%	1%	5%	2%	9%	2%	7%	1%	8%		1%	1%
d	1%	1%		3%	2%	1%	7%	5%	2%	3%		1%	1%
e	2%	4%	1%	3%	12%	5%	2%	9%	2%	4%		1%	1%
f	14%	1%	4%	2%	4%	5%	10%	6%	2%	6%		5%	1%
g				34%		1%		2%		1%		37%	1%
h	21%	1%	1%	1%	1%	2%	12%	12%	2%	25%			
i	11%	6%	2%	2%	3%	5%	7%	4%	2%	3%		3%	1%
j	13%	1%	6%	2%	1%	4%	8%	4%	3%	5%		7%	
k	7%		1%				7%	9%	2%	6%			
l	12%	3%	2%	1%	4%	5%	6%	6%	2%	7%		2%	1%
m	18%	1%	1%	1%	1%	2%	5%	17%	1%	11%		1%	
n	-	1%	4%	1%	5%	2%	2%	17%	7%	13%		1%	
o	15%	-	1%	2%	3%	1%	4%	6%	3%	6%		2%	1%
p	29%		-	2%	3%	1%	2%	2%	4%	1%		1%	1%
q		3%		-				12%					
r	19%	1%	3%	1%	-	9%	1%	9%	6%	10%		1%	1%
s	16%		11%	2%	4%	-	3%	9%	2%	4%		4%	2%
t	9%	2%	2%	1%	1%	6%	-	6%	2%	6%		1%	1%
u	22%	1%	1%	5%	1%	1%	4%	-	2%	9%		1%	
v	24%	1%	3%	1%	3%	1%	3%	12%	-	12%		2%	
w	29%	1%	1%	2%	5%	1%	3%	15%	4%	-		1%	
x											-		
y				23%				8%				-	
z	13%		20%			53%		7%					-

Table H-1: Letter Associations: confusion matrix for first letter

	a	b	c	d	e	f	g	h	i	j	k	l	m
a	-		1%	15%	23%	2%		1%				2%	2%
b	5%	-		5%	26%	5%							
c	1%		-	1%	79%			1%				3%	1%
d	3%			-	27%	1%		1%				9%	2%
e	1%		2%	21%	-	1%	1%	1%				6%	1%
f	5%			11%	25%	-		2%				2%	2%
g	1%			4%	1%		-					1%	
h	7%				11%			-				3%	8%
i	17%				9%	2%		2%	-			2%	2%
j									75%	-			
k	3%			14%	29%						-	15%	
l	1%		1%	11%	21%	1%	1%	1%	1%			-	4%
m	1%		1%	7%	9%		1%	1%				2%	-
n	1%		1%	6%	11%	1%		1%				4%	16%
o	1%				49%	1%						1%	4%
p						1%							6%
q	80%					20%							
r	1%		1%	5%	18%	1%	1%	2%				4%	5%
s	1%		1%	3%	7%	1%	3%	1%				2%	6%
t	1%		1%	16%	22%	1%	1%	1%				6%	5%
u	29%				24%	2%							5%
v													
w					6%			1%					5%
x					58%								
y	1%		1%	2%	3%	1%	39%	1%				3%	2%
z													

	n	o	p	q	r	s	t	u	v	w	x	y	z
a	15%	2%			15%	14%	3%	1%		3%	1%	1%	
b					16%		37%					5%	
c	4%	2%			5%	2%	2%				1%		
d	12%	1%			12%	4%	22%	1%		1%	1%	6%	
e	1%	14%			27%	4%	11%	1%		4%	2%	2%	
f	11%	5%			7%	8%	9%	2%		3%	1%	9%	
g	1%				2%	28%	2%					62%	
h	32%	7%			9%	11%	1%	1%		7%		3%	
i	9%	20%			11%	6%	15%			2%		5%	
j						5%	20%						
k	1%	19%			14%		5%					1%	
l	19%	4%			7%	6%	8%	1%		4%	1%	10%	
m	37%	2%			10%	12%	3%	3%		4%		9%	
n	-	2%			18%	15%	9%	4%		8%	1%	4%	
o	23%	-			11%	3%				4%	2%	1%	
p	61%	1%	-			9%						21%	
q				-									
r	35%	5%			-	7%	7%	1%		4%	1%	5%	
s	24%	2%			16%	-	9%	1%		1%	1%	25%	
t	18%	2%			5%	8%	-	1%		1%	1%	12%	
u	5%				17%	17%		-					
v									-				
w	33%	14%			35%		2%	2%		-	1%		
x	5%	11%			5%	5%	5%	5%		5%	-		
y	11%	1%			4%	27%	8%					-	
z													-

Table H-2: Letter Associations: confusion matrix for last letter

Table H-3 shows the percent correct recognition of the cues first letter and last letter using a range of constants for letter associations. Two constants have been tested. The first constant was used for letters found in candidate list, and the second for letters not found in candidate list. This difference was used to take account of the difference between letters which had already been suggested as possible candidates and those which had not. Letters, which were not found in the candidate list, could not be allocated a larger increase than those which had been found in the candidate list. Associated letters were increased by an amount proportional to the amount that the original letter was increased, to the degree of resemblance between the letters, and to a constant.

constant for letters found in candidate list	constant for letters not found in candidate list	percent first letter correct	percent last letter correct	average correct
0.0	0.0	37.6%	41.8%	39.7%
0.1	0.0	38.4%	41.8%	40.1%
0.1	0.1	38.4%	41.8%	40.1%
0.2	0.0	39.5%	42.5%	41.0%
0.2	0.1	39.5%	42.5%	41.0%
0.2	0.2	39.5%	42.5%	41.0%
0.3	0.0	39.5%	42.8%	41.2%
0.3	0.1	39.5%	42.8%	41.2%
0.3	0.2	39.5%	42.8%	41.2%
0.3	0.3	39.5%	42.8%	41.2%
0.4	0.0	39.5%	43.7%	41.6%
0.4	0.1	39.5%	43.7%	41.6%
0.4	0.2	39.5%	43.7%	41.6%
0.4	0.3	39.5%	43.7%	41.6%
0.4	0.4	39.5%	43.7%	41.6%
0.5	0.0	41.2%	45.1%	43.2%

constant for letters found in candidate list	constant for letters not found in candidate list	percent first letter correct	percent last letter correct	average correct
0.5	0.1	41.2%	45.1%	43.2%
0.5	0.2	41.2%	45.1%	43.2%
0.5	0.3	41.2%	45.1%	43.2%
0.5	0.4	41.2%	45.1%	43.2%
0.5	0.5	41.2%	45.1%	43.2%
0.6	0.0	42.0%	45.5%	43.7%
0.6	0.1	42.0%	45.5%	43.7%
0.6	0.2	42.0%	45.5%	43.7%
0.6	0.3	42.0%	45.5%	43.7%
0.6	0.4	42.0%	45.8%	43.9%
0.6	0.5	41.8%	45.8%	43.8%
0.6	0.6	41.7%	45.8%	43.7%
0.7	0.0	42.5%	47.3%	44.9%
0.7	0.1	42.5%	47.3%	44.9%
0.7	0.2	42.5%	47.3%	44.9%
0.7	0.3	42.5%	47.3%	44.9%
0.7	0.4	42.5%	48.6%	45.6%
0.7	0.5	42.3%	48.6%	45.5%
0.7	0.6	42.3%	48.6%	45.5%
0.7	0.7	41.8%	48.6%	45.2%
0.8	0.0	42.0%	48.1%	45.1%
0.8	0.1	42.0%	48.1%	45.1%
0.8	0.2	42.0%	48.1%	45.1%
0.8	0.3	42.0%	48.1%	45.1%
0.8	0.4	42.0%	48.1%	45.1%
0.8	0.5	42.0%	48.1%	45.1%
0.8	0.6	41.8%	48.1%	45.0%
0.8	0.7	41.8%	48.4%	45.1%
0.8	0.8	41.2%	48.4%	44.8%
0.9	0.0	41.8%	48.3%	45.1%
0.9	0.1	41.8%	48.3%	45.1%

constant for letters found in candidate list	constant for letters not found in candidate list	percent first letter correct	percent last letter correct	average correct
0.9	0.2	41.8%	48.3%	45.1%
0.9	0.3	41.8%	48.3%	45.1%
0.9	0.4	41.8%	48.3%	45.1%
0.9	0.5	41.8%	48.3%	45.1%
0.9	0.6	41.7%	48.3%	45.0%
0.9	0.7	41.5%	48.6%	45.1%
0.9	0.8	41.4%	48.6%	45.0%
0.9	0.9	40.4%	48.9%	44.6%
1.0	0.0	41.5%	48.1%	44.8%
1.0	0.1	41.5%	48.1%	44.8%
1.0	0.2	41.5%	48.1%	44.8%
1.0	0.3	41.5%	48.1%	44.8%
1.0	0.4	41.5%	48.1%	44.8%
1.0	0.5	41.5%	48.1%	44.8%
1.0	0.6	41.5%	48.1%	44.8%
1.0	0.7	41.5%	48.4%	45.0%
1.0	0.8	41.4%	48.4%	44.9%
1.0	0.9	40.7%	48.4%	44.6%

Table H-3: Letter Associations: percent correct for first and last letter using different constants: partial data set

Appendix I: Combining Methods for Cue Derivation

Results are shown here for combinations of different methods for obtaining word level cues (see section 5.2.7.1). These results show that best results are obtained when the combination of methods minimized dissimilarity, rather than maximized similarity. Dissimilarity is taken here to be a weak resemblance between the pattern of cues and the target. Similarity is taken to be a strong resemblance between the pattern of cues and the target.

Table I-1 shows the degree of resemblance between the target word and 18 different combinations of methods for deriving the cues. Results are given for the partial data set. The number of cues in each of the patterns that were identical to the cues of the target word was recorded. The table shows the percentage of patterns with 7 correct matches, with 6 or more correct matches, 5 or more correct matches, 1 correct match, 2 or less correct matches, and 3 or less correct matches.

combination id	7 correct matches	6 or more correct matches	5 or more correct matches	1 correct match	2 or less correct matches	3 or less correct matches
combination 1	3.6%	18.3%	42.8%	1.6%	10.0%	30.5%
combination 2	2.5%	19.8%	54.5%	0.0%	2.1%	15.7%
combination 3	3.1%	19.9%	54.7%	0.2%	2.5%	15.0%
combination 4	3.3%	22.4%	58.3%	0.2%	2.0%	16.0%
combination 5	4.1%	20.9%	45.8%	0.8%	8.9%	28.0%
combination 6	3.1%	22.7%	58.5%	0.0%	1.0%	13.8%
combination 7	4.1%	22.4%	58.5%	0.2%	1.3%	12.9%
combination 8	4.8%	24.7%	61.6%	0.2%	1.5%	13.0%
combination 9	4.1%	21.3%	45.6%	0.8%	8.6%	27.3%
combination 10	3.6%	22.6%	59.5%	0.0%	0.8%	14.2%
combination 11	4.6%	22.7%	58.6%	0.2%	1.5%	12.5%
combination 12	5.3%	24.7%	61.8%	0.2%	1.3%	12.7%
combination 13	3.3%	18.0%	42.3%	1.8%	10.4%	30.1%
combination 14	3.0%	21.3%	54.0%	0.0%	2.1%	15.8%
combination 15	3.3%	18.5%	43.0%	1.8%	10.4%	30.5%
combination 16	3.0%	21.4%	54.0%	0.0%	2.0%	15.8%
combination 17	3.3%	18.1%	43.0%	1.6%	10.2%	30.6%
combination 18	2.8%	21.3%	54.2%	0.0%	2.0%	15.8%

Table I-1: Degree of resemblance between the target word and 18 different combinations of methods for deriving the cues: partial data set

Table I-2 shows results obtained from applying the word level method using the 18 different combinations of methods. The version of the word level method used here is the one described in Chapter 6. Results are given for the partial data set. The table shows percent correct recognition for the target top ranked, percent correct recognition for target in the top 10, and the average position of the target in the list of candidates.

combination id	target top ranked	target in top 10	average position of target
combination 1	3.3%	15.2%	498.8
combination 2	4.9%	25.0%	200.3
combination 3	4.9%	24.2%	219.1
combination 4	4.6%	23.6%	223.6
combination 5	3.8%	16.8%	461.9
combination 6	6.1%	26.2%	183.9
combination 7	6.3%	26.4%	201.9
combination 8	6.3%	25.2%	203.9
combination 9	3.5%	16.1%	439.6
combination 10	6.1%	26.2%	178.6
combination 11	6.1%	26.0%	192.7
combination 12	6.4%	24.9%	190.6
combination 13	3.8%	16.3%	512.0
combination 14	4.6%	24.7%	213.0
combination 15	3.6%	15.2%	510.7
combination 16	4.6%	25.0%	211.9
combination 17	3.5%	15.2%	508.7
combination 18	4.5%	24.9%	210.9

Table I-2: The word level method applied to the 18 different combinations of methods: percent correct recognition of the target word plus average position of the target word: partial data set

Table I-3 shows indices of correlation between the results obtained from applying the word level method and the degree of resemblance between the target word and the patterns of cues. The Spearman rank correlation has been used because it measures the degree of agreement between two sets of ranks. The 18 different combinations of methods were firstly ranked according to the percentage of patterns with 7 correct matches, with 6 or more correct matches, 5 or more correct matches, 1 correct match, 2 or less correct matches, and 3 or less correct matches.

The top row of Table I-3 shows the correlation when the 18 different combinations of methods were ranked according to the percent correct recognition for the target top ranked. The middle row shows the correlation when the different combinations of methods were ranked according to the percent correct recognition for the target in the top 10. The bottom row shows the correlation when the 18 different combinations of methods were ranked according to the average position of the target in the list of candidates.

word level method	7 correct matches	6 or more correct matches	5 or more correct matches	1 correct match	2 or less correct matches	3 or less correct matches
target top ranked	r = 0.13, t = 0.52, df = 16	r = 0.74, t = 4.43, df = 16	r = 0.83, t = 5.88, df = 16	r = 0.80, t = 5.33, df = 16	r = 0.90, t = 8.20, df = 16	r = 0.88, t = 7.58, df = 16
target in top 10	r = 0.30, t = 1.25, df = 16	r = 0.80, t = 5.41, df = 16	r = 0.93, t = 9.94, df = 16	r = 0.66, t = 3.53, df = 16	r = 0.87, t = 6.92, df = 16	r = 0.93, t = 10.28, df = 16
average position of target	r = 0.16, t = 0.63, df = 16	r = 0.80, t = 5.29, df = 16	r = 0.87, t = 7.20, df = 16	r = 0.87, t = 7.13, df = 16	r = 0.95, t = 12.51, df = 16	r = 0.89, t = 7.92, df = 16

Table I-3: Indices of correlation between the word level method and the degree of resemblance between the target word and the patterns of cues: partial data set

The combinations of methods that minimized dissimilarity show a greater correlation to the accuracy of the word level method than combinations of methods that maximized similarity. These results suggest that the best combination of methods are those which minimized the likelihood of any of the values being incorrect when taken together.

Appendix J: Cue Derivation- Final Results

The tables in this appendix show the final results for cue derivation (see section 5.2.7.4). All results are for the partial data set.

The first ten tables Table J-1 to Table J-10 give the results for the different methods for limiting the number of candidates and for adjusting the confidence scores. Results are provided for the full range of the threshold tested for each method.

Results are shown for limiting the number of candidates by their score (score), by their rank (rank), by the difference between the confidence score of the top ranked candidate and the confidence scores of the other candidates (difference), and by the ratio between the confidence score of the top ranked candidate and the confidence scores of the other candidates (ratio). Table J-1 to Table J-4 present results for when weighting was not used. Table J-5 to Table J-8 present results for when weighting was used. Table J-9 and Table J-10 show results for weighting by the confidence scores raised to a set power (power) and for weighting by the normalized confidence scores raised to a set power (power plus normalization).

In Table J-1 to Table J-10 the results are given by the threshold used to limit the number of candidates or to adjust the confidence score of the candidates. The combination of alternatives that produced the most accurate results for a given cue has been used for all of the other methods.

threshold	length	ascender	descender	dot	cross	first	last	average
1	33.4%	80.4%	71.8%	74.0%	66.1%	35.1%	40.2%	57.3%
2	34.4%	80.4%	71.2%	79.6%	66.1%	35.1%	40.2%	58.1%
3	37.7%	82.0%	70.7%	80.1%	67.5%	34.8%	41.2%	59.1%
4	37.4%	82.2%	71.0%	79.1%	67.4%	35.4%	42.0%	59.2%
5	37.2%	81.1%	70.8%	80.2%	67.9%	37.2%	41.5%	59.4%
6	37.6%	80.9%	70.8%	79.9%	67.5%	37.4%	41.7%	59.4%
7	37.2%	80.7%	71.5%	79.1%	67.5%	38.2%	41.8%	59.4%
8	37.2%	80.9%	72.0%	79.7%	68.0%	38.4%	42.0%	59.8%
9	37.1%	80.9%	71.5%	80.1%	67.5%	37.9%	41.8%	59.5%
10	37.7%	80.9%	71.2%	80.1%	68.4%	37.6%	41.8%	59.7%
11	37.6%	80.6%	71.5%	79.9%	68.7%	37.6%	41.8%	59.7%
12	37.9%	80.6%	72.0%	80.4%	68.4%	37.2%	41.7%	59.7%
13	38.2%	80.4%	72.2%	80.9%	67.7%	36.6%	41.2%	59.6%
14	38.4%	80.6%	72.0%	79.9%	68.4%	36.1%	41.2%	59.5%
15	38.1%	80.7%	72.0%	80.6%	68.0%	36.2%	41.7%	59.6%
16	37.4%	81.1%	71.7%	80.7%	68.4%	36.1%	41.7%	59.6%
17	37.6%	80.9%	71.7%	80.1%	68.5%	36.1%	41.5%	59.5%
18	37.4%	80.9%	71.5%	80.2%	68.0%	36.4%	41.5%	59.4%
19	36.7%	80.9%	71.5%	80.2%	67.9%	36.2%	41.4%	59.3%
20	36.7%	80.7%	71.7%	80.2%	67.9%	36.2%	41.4%	59.3%
21	36.9%	80.4%	71.3%	80.2%	67.9%	35.9%	41.0%	59.1%
22	37.2%	80.6%	71.3%	80.1%	67.7%	35.7%	41.2%	59.1%
23	37.2%	80.4%	71.3%	80.2%	67.7%	35.6%	41.2%	59.1%
24	37.1%	80.7%	71.7%	80.2%	68.0%	35.7%	41.0%	59.2%
25	37.2%	80.6%	71.5%	80.2%	68.2%	35.9%	41.5%	59.3%
26	37.2%	80.7%	71.3%	80.2%	68.0%	36.1%	41.4%	59.3%

threshold	length	ascender	descender	dot	cross	first	last	average
27	37.4%	80.6%	71.3%	80.2%	67.9%	35.7%	41.8%	59.3%
28	37.2%	80.6%	71.3%	80.2%	68.4%	35.7%	41.8%	59.3%
29	37.1%	80.6%	71.0%	80.2%	68.0%	35.7%	41.7%	59.2%
30	37.1%	80.6%	71.0%	80.2%	68.2%	35.7%	41.5%	59.2%
31	37.1%	80.4%	71.0%	80.2%	68.2%	35.7%	41.7%	59.2%
32	36.9%	80.6%	71.0%	80.2%	68.4%	35.6%	41.5%	59.2%
33	36.9%	80.6%	71.0%	80.2%	68.2%	35.6%	41.5%	59.1%
34	37.1%	80.4%	71.0%	80.2%	68.4%	35.6%	41.8%	59.2%
35	37.1%	80.2%	71.2%	80.2%	68.2%	35.6%	42.0%	59.2%
36	36.9%	80.4%	71.2%	80.2%	68.4%	35.6%	42.0%	59.2%
37	36.7%	80.4%	71.2%	80.2%	68.4%	35.6%	42.0%	59.2%
38	36.7%	80.6%	71.2%	80.2%	68.4%	35.6%	42.2%	59.3%
39	36.7%	80.6%	71.2%	80.2%	68.2%	35.6%	42.3%	59.3%
40	36.9%	80.6%	71.2%	80.2%	68.2%	35.6%	42.5%	59.3%
41	36.9%	80.6%	71.2%	80.2%	68.4%	35.7%	42.5%	59.4%
42	36.9%	80.6%	71.0%	80.2%	68.5%	35.7%	42.3%	59.3%
43	36.9%	80.6%	71.0%	80.4%	68.4%	35.7%	42.2%	59.3%
44	36.9%	80.6%	70.8%	80.4%	68.4%	35.9%	41.8%	59.3%
45	37.1%	80.4%	70.8%	80.4%	68.4%	36.1%	42.2%	59.3%
46	37.1%	80.6%	70.8%	80.4%	68.4%	35.7%	42.3%	59.3%
47	37.1%	80.6%	70.8%	80.4%	68.5%	35.6%	42.3%	59.3%
48	36.9%	80.4%	70.8%	80.4%	68.5%	35.7%	42.2%	59.3%
49	36.7%	80.4%	70.8%	80.4%	68.4%	35.7%	42.0%	59.2%
50	36.7%	80.4%	71.0%	80.4%	68.4%	35.7%	42.0%	59.2%
51	36.9%	80.4%	70.8%	80.4%	68.5%	35.6%	42.3%	59.3%
52	36.9%	80.4%	70.8%	80.4%	68.5%	35.7%	42.3%	59.3%
53	36.6%	80.4%	70.8%	80.4%	68.5%	35.7%	42.3%	59.3%
54	36.6%	80.4%	70.7%	80.4%	68.5%	35.6%	42.5%	59.2%
55	36.6%	80.2%	70.5%	80.4%	68.4%	35.6%	42.3%	59.1%
56	36.6%	80.2%	70.7%	80.4%	68.5%	35.4%	42.3%	59.2%
57	36.4%	80.2%	71.0%	80.4%	68.4%	35.4%	42.3%	59.2%
58	36.6%	80.2%	71.0%	80.4%	68.4%	35.3%	42.3%	59.2%
59	36.7%	80.2%	70.7%	80.4%	68.4%	35.3%	42.3%	59.1%

threshold	length	ascender	descender	dot	cross	first	last	average
60	36.7%	80.2%	70.5%	80.4%	68.4%	35.3%	42.3%	59.1%
61	36.6%	80.2%	70.5%	80.4%	68.4%	35.3%	42.3%	59.1%
62	36.6%	80.2%	70.5%	80.4%	68.2%	35.3%	42.3%	59.1%
63	36.6%	80.2%	70.5%	80.4%	68.2%	35.1%	42.5%	59.1%
64	36.7%	80.2%	70.5%	80.4%	68.2%	35.1%	42.5%	59.1%
65	36.7%	80.2%	70.5%	80.4%	68.2%	35.1%	42.5%	59.1%
66	36.7%	80.2%	70.5%	80.4%	68.2%	35.1%	42.7%	59.1%
67	36.7%	80.2%	70.5%	80.4%	68.2%	35.4%	42.7%	59.2%
68	36.7%	80.2%	70.5%	80.4%	68.2%	35.4%	42.7%	59.2%
69	36.7%	80.2%	70.5%	80.4%	68.2%	35.4%	42.7%	59.2%
70	36.9%	80.2%	70.5%	80.4%	68.2%	35.4%	42.7%	59.2%
71	36.9%	80.2%	70.3%	80.4%	68.0%	35.1%	42.5%	59.1%
72	36.9%	80.2%	70.3%	80.4%	68.0%	35.1%	42.5%	59.1%
73	36.9%	80.2%	70.5%	80.4%	68.2%	35.1%	42.5%	59.1%
74	36.9%	80.2%	70.5%	80.4%	68.2%	35.1%	42.5%	59.1%
75	36.9%	80.2%	70.5%	80.4%	68.2%	35.1%	42.5%	59.1%
76	36.9%	80.2%	70.5%	80.4%	68.2%	35.1%	42.5%	59.1%
77	36.9%	80.2%	70.5%	80.4%	68.2%	35.1%	42.5%	59.1%
78	36.9%	80.2%	70.5%	80.4%	68.2%	35.1%	42.5%	59.1%
79	36.9%	80.2%	70.5%	80.4%	68.2%	35.1%	42.7%	59.1%
80	36.9%	80.2%	70.5%	80.4%	68.2%	35.1%	42.7%	59.1%
81	36.9%	80.2%	70.5%	80.4%	68.2%	35.1%	42.7%	59.1%
82	36.9%	80.1%	70.5%	80.4%	68.2%	35.1%	42.7%	59.1%
83	36.9%	80.1%	70.5%	80.4%	68.2%	35.1%	42.7%	59.1%
84	36.9%	80.1%	70.5%	80.4%	68.2%	35.1%	42.5%	59.1%
85	36.9%	80.1%	70.5%	80.4%	68.2%	35.1%	42.5%	59.1%
86	36.9%	80.1%	70.5%	80.4%	68.2%	35.3%	42.5%	59.1%
87	36.9%	80.1%	70.5%	80.4%	68.2%	35.3%	42.5%	59.1%
88	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
89	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
90	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
91	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
92	36.9%	80.1%	70.5%	80.4%	68.2%	35.3%	42.5%	59.1%

threshold	length	ascender	descender	dot	cross	first	last	average
93	36.9%	80.1%	70.5%	80.4%	68.2%	35.3%	42.5%	59.1%
94	36.9%	80.1%	70.5%	80.4%	68.2%	35.3%	42.5%	59.1%
95	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
96	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
97	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
98	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
99	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
100	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%

Table J-1: Percent correct of the cues: rank, without weighting, by threshold: partial data set

threshold	length	ascender	descender	dot	cross	first	last	average
100	33.4%	80.4%	71.8%	74.0%	66.1%	35.1%	40.2%	57.3%
99	33.6%	80.4%	71.7%	74.1%	66.2%	35.1%	40.2%	57.3%
98	33.6%	80.6%	71.5%	74.3%	66.1%	34.9%	40.2%	57.3%
97	33.8%	80.6%	71.5%	74.3%	66.1%	34.9%	40.2%	57.3%
96	33.9%	80.9%	71.3%	74.6%	66.1%	34.8%	40.5%	57.4%
95	34.8%	81.1%	71.3%	74.8%	66.1%	34.8%	40.9%	57.7%
94	34.6%	81.1%	71.3%	75.0%	66.4%	34.9%	40.9%	57.7%
93	34.9%	80.9%	71.3%	75.0%	65.9%	34.9%	41.2%	57.7%
92	35.1%	80.9%	71.5%	75.0%	65.9%	35.1%	41.4%	57.8%
91	35.3%	80.7%	71.7%	74.8%	66.1%	35.1%	41.4%	57.8%
90	35.4%	80.7%	71.5%	75.1%	66.1%	34.9%	41.5%	57.9%
89	35.4%	80.7%	71.5%	75.1%	66.1%	35.1%	41.5%	57.9%
88	35.1%	80.7%	71.2%	75.3%	65.7%	34.9%	41.4%	57.8%
87	35.3%	80.7%	71.3%	75.5%	65.9%	35.1%	41.5%	57.9%
86	34.9%	80.9%	71.0%	75.5%	66.1%	35.1%	41.4%	57.8%
85	35.3%	80.9%	70.8%	75.8%	66.4%	35.3%	41.4%	58.0%
84	35.7%	80.9%	70.8%	75.8%	66.4%	35.1%	41.4%	58.0%
83	35.4%	80.9%	71.3%	75.8%	66.4%	34.9%	41.4%	58.0%
82	35.7%	81.1%	70.8%	75.8%	66.2%	34.9%	41.4%	58.0%
81	35.9%	80.9%	70.7%	75.8%	66.1%	35.1%	40.9%	57.9%

threshold	length	ascender	descender	dot	cross	first	last	average
80	36.4%	80.4%	70.8%	75.5%	66.1%	35.4%	40.9%	57.9%
79	36.7%	80.6%	71.0%	75.8%	65.7%	35.6%	41.2%	58.1%
78	36.4%	80.6%	71.0%	76.1%	66.2%	35.7%	40.9%	58.1%
77	36.2%	80.7%	71.7%	76.1%	65.9%	35.7%	40.9%	58.2%
76	36.4%	80.6%	72.0%	75.5%	65.7%	35.4%	41.0%	58.1%
75	36.6%	80.6%	71.8%	75.8%	66.2%	35.6%	41.0%	58.2%
74	36.9%	80.6%	72.0%	76.4%	66.6%	35.3%	40.9%	58.4%
73	36.9%	80.9%	71.7%	76.4%	66.2%	35.4%	41.0%	58.4%
72	36.4%	81.1%	71.7%	75.9%	66.7%	34.9%	41.4%	58.3%
71	36.7%	81.1%	71.8%	75.9%	66.7%	35.1%	41.4%	58.4%
70	37.1%	81.1%	72.2%	75.8%	66.4%	34.9%	41.7%	58.4%
69	36.6%	81.1%	72.2%	76.1%	66.7%	34.9%	41.7%	58.5%
68	36.4%	80.9%	72.2%	75.6%	66.7%	34.8%	41.7%	58.3%
67	36.2%	81.1%	72.7%	75.6%	66.7%	34.8%	41.4%	58.3%
66	36.4%	80.7%	72.5%	75.9%	66.7%	34.6%	41.5%	58.3%
65	36.4%	80.6%	72.3%	76.1%	66.9%	34.8%	42.0%	58.4%
64	36.7%	80.2%	72.3%	76.8%	67.4%	34.8%	42.2%	58.6%
63	36.2%	80.4%	72.2%	76.6%	67.2%	35.1%	42.2%	58.6%
62	36.7%	80.2%	72.2%	76.6%	67.4%	34.8%	42.3%	58.6%
61	36.9%	80.2%	71.8%	76.4%	67.5%	35.3%	42.3%	58.6%
60	37.1%	79.9%	71.7%	76.1%	67.2%	35.1%	42.5%	58.5%
59	37.7%	80.4%	72.0%	76.6%	67.1%	35.6%	42.7%	58.9%
58	37.4%	80.2%	72.2%	76.6%	66.7%	35.9%	42.3%	58.8%
57	37.4%	80.7%	72.2%	76.3%	67.1%	36.4%	41.7%	58.8%
56	37.2%	81.1%	72.2%	76.6%	67.7%	36.1%	41.5%	58.9%
55	36.9%	81.1%	72.2%	76.9%	67.4%	36.1%	41.5%	58.9%
54	36.6%	80.9%	71.7%	76.6%	67.5%	35.7%	41.4%	58.6%
53	37.1%	80.6%	71.5%	76.9%	67.2%	35.7%	41.5%	58.6%
52	37.4%	80.9%	71.2%	77.3%	67.7%	35.6%	41.5%	58.8%
51	37.7%	80.9%	71.3%	77.4%	67.5%	36.1%	41.7%	59.0%
50	38.2%	80.9%	71.2%	77.6%	67.5%	36.4%	42.0%	59.1%
49	38.2%	80.6%	71.0%	78.3%	67.7%	37.4%	41.4%	59.2%
48	38.4%	80.6%	71.0%	77.6%	66.9%	37.1%	41.5%	59.0%

threshold	length	ascender	descender	dot	cross	first	last	average
47	38.1%	80.9%	71.2%	77.8%	66.9%	36.9%	41.8%	59.1%
46	38.4%	80.9%	71.0%	77.8%	67.1%	36.6%	41.8%	59.1%
45	38.4%	80.6%	70.8%	77.9%	67.1%	35.7%	41.5%	58.9%
44	38.2%	80.9%	71.0%	77.8%	67.7%	35.3%	41.7%	58.9%
43	37.9%	80.6%	70.5%	77.9%	67.9%	35.4%	41.5%	58.8%
42	38.2%	80.4%	71.3%	78.1%	68.2%	35.4%	41.8%	59.1%
41	37.9%	80.2%	71.5%	78.1%	69.5%	35.6%	41.8%	59.2%
40	38.2%	80.1%	71.2%	78.4%	69.4%	35.6%	42.5%	59.3%
39	38.1%	80.1%	71.7%	78.9%	68.7%	35.7%	42.5%	59.4%
38	38.2%	80.1%	72.3%	79.6%	68.9%	35.7%	42.5%	59.6%
37	37.7%	80.1%	71.8%	79.6%	69.0%	35.7%	42.5%	59.5%
36	37.4%	80.1%	71.8%	79.6%	68.7%	35.7%	42.7%	59.4%
35	37.4%	80.4%	71.7%	79.6%	69.5%	36.1%	42.5%	59.6%
34	37.9%	80.4%	71.7%	79.2%	69.4%	35.7%	42.8%	59.6%
33	37.2%	80.4%	71.2%	79.2%	69.0%	36.1%	43.0%	59.4%
32	37.1%	80.2%	71.5%	79.6%	68.7%	36.2%	43.2%	59.5%
31	36.7%	80.2%	71.5%	80.1%	69.0%	35.9%	42.8%	59.5%
30	37.2%	80.2%	72.2%	80.1%	69.0%	35.9%	43.2%	59.7%
29	37.4%	80.2%	72.8%	79.9%	68.7%	35.9%	42.7%	59.7%
28	37.7%	80.1%	72.2%	80.2%	68.5%	35.7%	42.8%	59.6%
27	37.2%	79.9%	71.7%	80.4%	68.0%	35.4%	42.3%	59.3%
26	37.6%	79.9%	71.8%	80.2%	68.4%	35.6%	42.2%	59.4%
25	37.9%	79.9%	71.0%	80.2%	68.7%	35.6%	42.3%	59.4%
24	37.9%	79.9%	70.8%	80.2%	68.2%	34.9%	42.2%	59.2%
23	37.2%	79.9%	70.8%	80.2%	67.9%	35.1%	42.0%	59.0%
22	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
21	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
20	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
19	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
18	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
17	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
16	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
15	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%

threshold	length	ascender	descender	dot	cross	first	last	average
14	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
13	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
12	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
11	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
10	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
9	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
8	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
7	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
6	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
5	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
4	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
3	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
2	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
1	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%

Table J-2: Percent correct of the cues: score, without weighting, by threshold: partial data set

threshold	length	ascender	descender	dot	cross	first	last	average
0	33.4%	80.4%	71.8%	74.0%	66.1%	35.1%	40.2%	57.3%
1	34.6%	80.4%	72.3%	74.3%	66.4%	35.1%	40.0%	57.6%
2	34.3%	80.6%	72.5%	75.3%	67.1%	34.9%	40.2%	57.8%
3	34.9%	80.9%	71.7%	75.8%	66.9%	35.1%	40.0%	57.9%
4	35.6%	81.1%	71.7%	75.9%	66.4%	34.8%	40.7%	58.0%
5	35.9%	80.7%	71.7%	77.1%	66.6%	34.9%	41.2%	58.3%
6	36.7%	80.7%	71.5%	77.1%	66.6%	35.1%	40.7%	58.3%
7	37.1%	80.7%	71.0%	77.6%	66.4%	35.6%	41.7%	58.6%
8	37.6%	80.9%	71.2%	77.8%	66.6%	35.1%	41.8%	58.7%
9	37.2%	80.7%	71.2%	77.3%	66.6%	35.7%	41.5%	58.6%
10	37.6%	80.7%	71.5%	77.9%	66.7%	35.7%	42.0%	58.9%
11	38.1%	81.2%	71.5%	78.1%	66.4%	36.1%	42.2%	59.1%
12	37.2%	81.1%	71.2%	78.1%	66.2%	36.2%	41.5%	58.8%
13	37.6%	81.2%	71.5%	78.6%	66.6%	36.4%	42.2%	59.1%
14	37.4%	81.4%	71.2%	78.4%	66.9%	36.6%	41.8%	59.1%
15	38.2%	81.4%	71.2%	79.1%	67.1%	36.7%	42.2%	59.4%
16	38.4%	81.4%	71.8%	79.4%	67.2%	37.2%	42.0%	59.6%
17	37.9%	81.2%	71.5%	79.6%	67.1%	37.1%	41.7%	59.4%
18	37.7%	81.4%	71.8%	79.6%	66.9%	37.4%	41.5%	59.5%
19	38.2%	81.2%	71.8%	79.7%	66.6%	37.2%	40.7%	59.4%
20	38.7%	80.6%	71.7%	79.9%	66.6%	37.1%	40.5%	59.3%

threshold	length	ascender	descender	dot	cross	first	last	average
21	39.4%	81.1%	71.8%	79.9%	65.9%	37.2%	40.4%	59.4%
22	39.4%	80.9%	72.0%	80.4%	65.9%	37.6%	40.0%	59.4%
23	38.9%	80.9%	72.3%	80.2%	65.1%	37.7%	40.2%	59.3%
24	38.7%	80.6%	72.3%	79.4%	65.2%	37.2%	40.4%	59.1%
25	39.0%	80.6%	72.0%	80.1%	65.9%	37.6%	40.5%	59.4%
26	38.7%	80.7%	71.8%	80.2%	66.9%	37.4%	40.7%	59.5%
27	38.7%	81.1%	71.7%	80.1%	66.6%	37.2%	40.4%	59.4%
28	38.4%	81.2%	71.5%	79.6%	66.9%	36.7%	40.9%	59.3%
29	38.7%	81.4%	71.8%	79.1%	67.4%	36.9%	41.0%	59.5%
30	38.9%	81.5%	71.5%	79.2%	66.9%	36.6%	41.0%	59.4%
31	38.6%	81.9%	71.7%	79.1%	66.9%	36.6%	41.4%	59.4%
32	38.1%	81.9%	71.5%	78.6%	67.5%	36.6%	41.2%	59.3%
33	38.4%	82.4%	71.2%	78.7%	68.0%	36.9%	41.2%	59.5%
34	38.4%	82.2%	71.2%	79.6%	68.0%	36.9%	41.4%	59.7%
35	38.4%	82.0%	70.8%	79.7%	68.5%	36.9%	41.5%	59.7%
36	38.2%	81.7%	70.8%	79.7%	68.7%	36.9%	41.5%	59.7%
37	38.9%	81.9%	70.7%	79.2%	68.9%	37.1%	41.5%	59.7%
38	39.2%	81.9%	70.5%	79.1%	69.2%	36.7%	41.4%	59.7%
39	38.9%	81.9%	70.5%	78.9%	69.4%	37.2%	41.4%	59.7%
40	38.9%	81.5%	70.7%	78.9%	68.9%	37.1%	41.7%	59.7%
41	38.7%	81.9%	71.0%	79.1%	68.5%	37.1%	41.7%	59.7%
42	38.9%	81.7%	71.5%	79.2%	68.4%	37.2%	42.2%	59.9%
43	38.1%	82.4%	71.5%	79.2%	68.7%	37.4%	42.3%	59.9%
44	38.1%	82.2%	71.7%	79.6%	68.7%	36.7%	42.3%	59.9%
45	37.2%	82.0%	71.3%	79.9%	68.4%	36.7%	42.2%	59.7%
46	37.6%	82.0%	71.5%	79.9%	68.5%	36.1%	42.2%	59.7%
47	37.2%	81.9%	70.7%	79.9%	68.0%	36.1%	42.0%	59.4%
48	37.4%	82.0%	70.2%	79.9%	68.0%	35.9%	42.0%	59.4%
49	37.1%	82.0%	70.2%	80.2%	68.4%	36.6%	42.0%	59.5%
50	37.4%	81.9%	70.0%	80.4%	68.0%	36.4%	42.2%	59.5%
51	37.4%	81.2%	69.9%	80.7%	68.0%	36.7%	42.2%	59.4%
52	37.2%	81.4%	69.7%	80.4%	67.9%	36.4%	42.3%	59.3%
53	37.1%	81.4%	70.2%	80.6%	67.7%	36.4%	42.7%	59.4%

threshold	length	ascender	descender	dot	cross	first	last	average
54	36.9%	81.2%	70.0%	80.6%	67.9%	36.1%	42.7%	59.3%
55	37.2%	80.9%	70.0%	80.4%	67.7%	35.6%	42.3%	59.2%
56	36.9%	81.1%	69.5%	79.9%	67.7%	35.1%	42.2%	58.9%
57	37.1%	80.7%	69.4%	80.1%	67.9%	35.1%	42.3%	58.9%
58	36.9%	80.6%	69.9%	80.1%	67.4%	34.9%	42.3%	58.9%
59	36.9%	80.4%	70.0%	80.1%	68.2%	35.3%	42.3%	59.0%
60	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
61	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
62	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
63	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
64	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
65	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
66	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
67	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
68	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
69	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
70	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
71	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
72	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
73	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
74	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
75	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
76	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
77	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
78	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
79	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
80	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
81	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
82	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
83	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
84	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
85	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
86	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%

threshold	length	ascender	descender	dot	cross	first	last	average
87	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
88	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
89	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
90	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
91	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
92	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
93	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
94	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
95	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
96	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
97	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
98	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
99	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%

Table J-3: Percent correct of the cues: difference, without weighting, by threshold: partial data set

threshold	length	ascender	descender	dot	cross	first	last	average
99	33.4%	80.4%	71.8%	74.0%	66.1%	35.1%	40.2%	57.3%
98	34.1%	80.4%	72.2%	74.3%	66.2%	35.1%	40.2%	57.5%
97	34.3%	80.6%	72.0%	74.6%	66.6%	34.9%	40.2%	57.6%
96	34.8%	80.6%	72.3%	74.8%	66.7%	34.9%	40.2%	57.8%
95	34.9%	80.9%	71.8%	75.3%	66.6%	34.8%	40.4%	57.8%
94	35.3%	80.9%	72.0%	75.6%	66.4%	34.9%	40.7%	58.0%
93	35.9%	80.7%	71.7%	76.3%	66.7%	35.1%	40.9%	58.2%
92	36.2%	80.7%	71.3%	76.8%	66.1%	35.1%	41.4%	58.2%
91	36.7%	81.1%	71.5%	77.1%	66.4%	35.1%	41.8%	58.5%
90	36.6%	80.9%	71.7%	77.1%	66.2%	35.1%	41.8%	58.5%
89	36.7%	80.7%	71.2%	77.4%	66.4%	34.9%	41.8%	58.5%
88	36.9%	80.7%	71.0%	77.6%	66.7%	35.3%	41.7%	58.6%
87	36.2%	80.7%	71.0%	77.8%	66.6%	35.3%	41.5%	58.4%
86	36.6%	80.9%	71.2%	77.8%	66.6%	35.4%	41.4%	58.5%
85	36.4%	81.1%	70.8%	77.4%	66.7%	35.7%	41.2%	58.5%
84	37.9%	81.1%	71.0%	77.6%	67.1%	36.2%	41.4%	58.9%
83	38.2%	81.2%	71.3%	77.6%	67.7%	36.2%	41.5%	59.1%
82	37.6%	80.9%	71.5%	77.6%	67.9%	36.1%	41.7%	59.0%
81	37.6%	81.2%	71.8%	78.1%	67.5%	36.1%	41.7%	59.1%
80	38.2%	81.1%	71.3%	78.3%	67.4%	36.1%	40.9%	59.0%

threshold	length	ascender	descender	dot	cross	first	last	average
79	39.2%	80.6%	71.7%	78.4%	67.2%	36.4%	40.7%	59.2%
78	38.7%	80.7%	72.2%	78.4%	66.7%	36.4%	41.5%	59.2%
77	38.9%	80.6%	72.5%	78.9%	67.1%	36.9%	41.0%	59.4%
76	38.7%	80.6%	73.6%	79.2%	67.2%	36.7%	40.9%	59.6%
75	38.9%	80.6%	73.6%	78.7%	67.1%	36.6%	40.9%	59.5%
74	39.2%	80.4%	73.0%	79.4%	67.7%	37.1%	40.7%	59.6%
73	39.2%	80.7%	72.7%	80.1%	67.5%	37.6%	41.4%	59.9%
72	39.4%	80.9%	71.8%	79.9%	67.4%	37.9%	41.2%	59.8%
71	38.7%	80.9%	71.7%	79.2%	67.4%	37.2%	41.5%	59.5%
70	38.7%	81.2%	72.0%	79.1%	67.5%	37.4%	41.5%	59.6%
69	39.2%	81.4%	72.2%	79.1%	67.1%	36.7%	41.7%	59.6%
68	39.2%	81.4%	72.5%	79.6%	67.4%	36.6%	41.5%	59.7%
67	39.2%	81.4%	72.3%	78.9%	67.2%	37.1%	41.7%	59.7%
66	39.0%	81.9%	72.2%	79.1%	66.7%	36.6%	41.8%	59.6%
65	38.7%	81.9%	72.0%	79.1%	66.9%	36.4%	41.7%	59.5%
64	38.4%	81.9%	72.0%	79.1%	67.2%	36.9%	41.8%	59.6%
63	38.4%	81.5%	71.8%	79.2%	67.7%	37.1%	42.3%	59.7%
62	38.1%	81.9%	71.5%	79.1%	67.5%	37.1%	42.2%	59.6%
61	38.4%	81.9%	71.3%	79.1%	67.9%	36.7%	41.8%	59.6%
60	38.6%	81.9%	71.3%	79.2%	67.7%	37.1%	41.8%	59.7%
59	38.9%	81.4%	71.2%	79.4%	67.2%	37.1%	42.0%	59.6%
58	39.0%	81.5%	71.2%	79.6%	67.2%	36.7%	41.8%	59.6%
57	39.4%	81.5%	70.7%	79.4%	67.5%	37.2%	42.3%	59.7%
56	39.4%	81.7%	71.0%	79.1%	67.7%	37.2%	42.0%	59.7%
55	39.2%	81.5%	70.7%	79.4%	67.7%	36.9%	42.2%	59.7%
54	38.2%	81.5%	70.7%	79.7%	67.4%	37.2%	41.7%	59.5%
53	38.6%	81.7%	70.7%	79.7%	68.0%	37.1%	41.7%	59.6%
52	38.1%	81.5%	70.5%	79.9%	67.9%	37.1%	41.7%	59.5%
51	37.7%	81.7%	70.2%	79.2%	68.2%	36.7%	42.0%	59.4%
50	38.2%	81.9%	70.5%	79.6%	68.0%	37.1%	42.0%	59.6%
49	38.4%	81.7%	70.3%	79.9%	67.5%	36.9%	42.3%	59.6%
48	37.6%	81.1%	70.2%	80.1%	67.5%	37.7%	42.2%	59.5%
47	37.2%	81.2%	69.9%	79.7%	67.5%	37.6%	42.3%	59.4%

threshold	length	ascender	descender	dot	cross	first	last	average
46	37.1%	81.4%	70.2%	80.2%	67.7%	37.4%	42.3%	59.5%
45	37.4%	81.2%	70.2%	80.2%	67.5%	37.2%	42.0%	59.4%
44	37.7%	81.1%	70.5%	80.1%	67.7%	36.7%	42.2%	59.4%
43	37.2%	81.2%	70.7%	79.6%	67.9%	35.9%	41.8%	59.2%
42	37.4%	80.9%	70.5%	79.7%	67.7%	35.9%	41.8%	59.1%
41	37.7%	80.6%	70.7%	79.9%	67.5%	35.4%	42.0%	59.1%
40	37.7%	80.4%	71.0%	80.1%	68.4%	35.6%	42.0%	59.3%
39	37.7%	80.2%	71.3%	80.4%	68.2%	35.7%	42.2%	59.4%
38	37.4%	80.2%	71.2%	80.4%	68.5%	35.9%	42.5%	59.4%
37	37.4%	80.2%	71.2%	80.4%	68.2%	35.9%	42.7%	59.4%
36	37.4%	80.2%	71.0%	80.4%	68.0%	35.9%	42.2%	59.3%
35	37.6%	80.2%	70.8%	80.6%	67.9%	35.9%	42.2%	59.3%
34	37.6%	80.2%	70.7%	80.6%	68.2%	35.9%	42.2%	59.3%
33	37.6%	80.1%	70.5%	80.6%	68.4%	35.4%	42.0%	59.2%
32	37.6%	80.1%	70.3%	80.4%	68.5%	35.4%	42.5%	59.3%
31	37.6%	80.1%	70.5%	80.4%	68.5%	35.3%	42.5%	59.3%
30	37.2%	80.1%	70.5%	80.4%	68.5%	35.3%	42.3%	59.2%
29	37.1%	80.1%	70.5%	80.4%	68.5%	35.3%	42.5%	59.2%
28	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
27	36.9%	80.1%	70.5%	80.4%	68.2%	35.3%	42.5%	59.1%
26	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
25	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
24	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
23	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
22	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
21	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
20	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
19	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
18	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
17	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
16	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
15	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
14	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%

threshold	length	ascender	descender	dot	cross	first	last	average
13	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
12	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
11	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
10	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
9	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
8	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
7	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
6	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
5	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
4	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
3	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
2	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%
1	36.9%	80.1%	70.5%	80.4%	68.4%	35.3%	42.5%	59.1%

Table J-4: Percent correct of the cues: ratio, without weighting, by threshold: partial data set

threshold	length	ascender	descender	dot	cross	first	last	average
1	33.4%	80.4%	71.8%	74.0%	66.1%	35.1%	41.2%	57.4%
2	34.4%	80.4%	71.5%	79.7%	65.6%	38.7%	45.3%	59.4%
3	37.6%	82.0%	70.8%	79.4%	67.7%	38.2%	45.0%	60.1%
4	38.9%	82.5%	71.3%	79.6%	67.5%	40.2%	46.8%	61.0%
5	37.7%	81.5%	71.3%	80.2%	67.9%	41.7%	47.0%	61.0%
6	37.7%	81.5%	71.2%	79.4%	67.4%	42.0%	47.3%	60.9%
7	38.2%	81.4%	71.7%	78.6%	68.0%	41.7%	47.4%	61.0%
8	36.9%	81.2%	72.2%	79.4%	68.4%	42.0%	48.3%	61.2%
9	37.7%	81.4%	72.3%	79.9%	67.9%	42.0%	47.8%	61.3%
10	38.6%	81.2%	72.2%	79.6%	68.5%	42.2%	47.4%	61.4%
11	38.1%	81.2%	72.3%	79.6%	69.2%	42.2%	47.9%	61.5%
12	38.1%	81.4%	73.0%	79.9%	69.4%	42.2%	47.4%	61.6%
13	37.9%	81.1%	73.1%	79.7%	68.4%	42.5%	47.8%	61.5%
14	38.4%	81.4%	72.7%	79.7%	69.2%	42.2%	47.9%	61.6%
15	37.9%	81.4%	73.0%	80.2%	69.0%	42.5%	47.4%	61.6%
16	38.4%	81.5%	72.3%	80.2%	69.0%	42.7%	47.6%	61.7%
17	38.4%	81.7%	72.3%	80.1%	68.9%	42.7%	47.4%	61.6%
18	38.4%	81.5%	72.3%	80.1%	69.0%	42.3%	47.8%	61.6%
19	38.2%	81.5%	72.7%	80.1%	68.5%	42.2%	47.3%	61.5%
20	38.2%	81.2%	72.7%	79.9%	68.7%	42.2%	47.0%	61.4%

threshold	length	ascender	descender	dot	cross	first	last	average
21	37.7%	81.4%	72.0%	79.7%	68.5%	42.5%	46.8%	61.2%
22	37.9%	81.2%	72.0%	79.7%	68.5%	42.2%	47.1%	61.2%
23	38.7%	81.4%	72.2%	79.9%	68.4%	42.3%	47.3%	61.4%
24	38.4%	81.4%	72.2%	80.1%	68.4%	42.3%	47.3%	61.4%
25	38.4%	81.2%	72.2%	80.1%	68.4%	41.8%	47.3%	61.3%
26	38.4%	81.1%	71.8%	80.1%	68.5%	42.0%	47.4%	61.3%
27	38.7%	81.2%	72.0%	80.1%	68.4%	42.0%	47.4%	61.4%
28	38.4%	81.2%	71.8%	79.9%	68.9%	42.0%	47.4%	61.4%
29	38.4%	81.2%	71.7%	79.9%	68.7%	41.8%	47.4%	61.3%
30	38.2%	81.2%	71.5%	79.9%	68.5%	41.8%	47.4%	61.2%
31	38.4%	81.2%	71.5%	79.9%	68.5%	41.8%	47.1%	61.2%
32	38.6%	81.2%	71.7%	79.9%	68.4%	42.2%	47.1%	61.3%
33	38.7%	81.2%	71.5%	79.9%	68.5%	42.2%	47.3%	61.3%
34	38.6%	81.1%	71.5%	79.9%	68.5%	42.2%	47.4%	61.3%
35	38.6%	81.2%	71.7%	79.9%	68.7%	42.0%	47.3%	61.3%
36	38.6%	81.2%	71.7%	79.9%	68.7%	42.0%	47.3%	61.3%
37	38.6%	81.2%	71.7%	79.9%	68.7%	41.8%	47.4%	61.3%
38	38.4%	81.2%	71.7%	79.9%	68.7%	41.7%	47.3%	61.3%
39	38.2%	81.4%	71.7%	79.9%	68.7%	41.7%	47.3%	61.3%
40	38.2%	81.4%	71.7%	79.9%	68.5%	41.8%	47.4%	61.3%
41	38.1%	81.4%	71.7%	80.1%	68.5%	41.8%	47.3%	61.3%
42	38.1%	81.4%	71.7%	80.1%	68.5%	41.8%	47.0%	61.2%
43	38.1%	81.4%	71.7%	80.1%	68.5%	41.8%	46.8%	61.2%
44	38.2%	81.2%	71.7%	80.1%	68.4%	41.8%	46.8%	61.2%
45	38.4%	81.2%	71.7%	80.1%	68.5%	42.2%	46.8%	61.3%
46	38.6%	81.2%	71.7%	80.1%	68.7%	42.2%	46.8%	61.3%
47	38.6%	81.2%	71.7%	80.1%	68.7%	42.2%	47.0%	61.3%
48	38.6%	81.1%	71.7%	80.1%	68.7%	42.2%	47.0%	61.3%
49	38.6%	81.1%	71.7%	80.1%	68.7%	42.2%	47.0%	61.3%
50	38.6%	81.1%	71.7%	80.1%	68.9%	42.0%	47.0%	61.3%
51	38.6%	81.2%	71.7%	80.1%	68.7%	42.0%	46.6%	61.3%
52	38.6%	81.1%	71.7%	80.1%	68.9%	42.0%	46.6%	61.3%
53	38.6%	81.2%	71.7%	80.1%	68.9%	42.0%	46.6%	61.3%

threshold	length	ascender	descender	dot	cross	first	last	average
54	38.6%	81.2%	71.5%	80.1%	68.9%	41.8%	46.6%	61.2%
55	38.6%	81.2%	71.5%	80.1%	68.9%	41.8%	46.6%	61.2%
56	38.6%	81.2%	71.5%	80.1%	68.7%	41.8%	46.6%	61.2%
57	38.6%	81.2%	71.5%	80.1%	68.9%	41.8%	46.6%	61.2%
58	38.6%	81.2%	71.5%	80.1%	68.9%	41.8%	46.6%	61.2%
59	38.7%	81.2%	71.5%	80.1%	68.9%	41.8%	46.6%	61.3%
60	38.7%	81.2%	71.5%	80.1%	68.9%	41.8%	46.6%	61.3%
61	38.7%	81.2%	71.5%	80.1%	68.9%	41.8%	46.6%	61.3%
62	38.7%	81.2%	71.3%	80.1%	68.7%	41.8%	46.8%	61.2%
63	38.6%	81.2%	71.3%	80.1%	68.7%	41.8%	46.8%	61.2%
64	38.4%	81.2%	71.5%	80.1%	68.7%	41.8%	46.8%	61.2%
65	38.4%	81.2%	71.5%	80.1%	68.7%	41.8%	46.8%	61.2%
66	38.6%	81.2%	71.5%	80.1%	68.7%	41.8%	46.8%	61.2%
67	38.6%	81.1%	71.5%	80.1%	68.7%	41.8%	46.8%	61.2%
68	38.6%	81.1%	71.5%	80.1%	68.7%	41.8%	46.8%	61.2%
69	38.6%	81.1%	71.5%	80.1%	68.7%	41.8%	46.8%	61.2%
70	38.6%	81.1%	71.5%	80.1%	68.7%	41.8%	46.8%	61.2%
71	38.6%	81.1%	71.5%	80.1%	68.7%	41.8%	46.8%	61.2%
72	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.8%	61.2%
73	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.8%	61.2%
74	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.8%	61.2%
75	38.6%	80.9%	71.7%	80.1%	68.7%	41.8%	46.8%	61.2%
76	38.6%	80.9%	71.7%	80.1%	68.7%	41.8%	46.6%	61.2%
77	38.6%	80.9%	71.7%	80.1%	68.7%	41.8%	46.6%	61.2%
78	38.6%	80.9%	71.7%	80.1%	68.7%	41.8%	46.6%	61.2%
79	38.6%	80.9%	71.5%	80.1%	68.7%	41.7%	46.6%	61.1%
80	38.6%	80.9%	71.5%	80.1%	68.7%	41.7%	46.6%	61.1%
81	38.6%	80.9%	71.5%	80.1%	68.7%	41.7%	46.6%	61.1%
82	38.6%	80.9%	71.5%	80.1%	68.7%	41.7%	46.5%	61.1%
83	38.6%	80.9%	71.5%	80.1%	68.7%	41.7%	46.5%	61.1%
84	38.6%	80.9%	71.5%	80.1%	68.7%	41.7%	46.5%	61.1%
85	38.6%	80.9%	71.5%	80.1%	68.7%	41.7%	46.5%	61.1%
86	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%

threshold	length	ascender	descender	dot	cross	first	last	average
87	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
88	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
89	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
90	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
91	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
92	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
93	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
94	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
95	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
96	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
97	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
98	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
99	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
100	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%

Table J-5: Percent correct of the cues: rank, with weighting, by threshold: partial data set

threshold	length	ascender	descender	dot	cross	first	last	average
100	33.4%	80.4%	71.8%	74.0%	66.1%	35.1%	41.2%	57.4%
99	33.4%	80.4%	71.8%	74.1%	66.1%	35.4%	41.7%	57.6%
98	33.6%	80.6%	71.8%	74.3%	66.1%	35.3%	41.8%	57.6%
97	33.8%	80.6%	71.8%	74.3%	66.1%	35.4%	42.0%	57.7%
96	33.9%	80.9%	71.7%	74.6%	66.1%	35.3%	42.0%	57.8%
95	34.8%	80.7%	71.7%	74.8%	66.1%	35.3%	42.7%	58.0%
94	34.6%	80.7%	71.7%	75.0%	66.1%	35.6%	42.7%	58.0%
93	34.9%	80.7%	71.8%	75.0%	66.1%	35.7%	42.7%	58.1%
92	35.1%	80.9%	71.8%	75.0%	66.2%	35.9%	43.0%	58.3%
91	35.3%	80.7%	71.8%	74.8%	66.4%	36.1%	43.2%	58.3%
90	35.4%	80.7%	71.8%	75.1%	66.4%	35.6%	43.8%	58.4%
89	35.3%	80.7%	71.7%	75.1%	66.4%	35.6%	43.5%	58.3%
88	35.1%	80.7%	71.5%	75.3%	66.1%	35.9%	43.7%	58.3%
87	35.1%	80.7%	71.3%	75.5%	66.1%	36.6%	43.5%	58.4%
86	34.9%	80.9%	71.2%	75.3%	66.2%	36.6%	43.3%	58.3%
85	35.3%	80.9%	71.2%	75.3%	66.4%	36.7%	43.2%	58.4%
84	35.7%	80.7%	71.2%	75.5%	66.7%	36.7%	43.3%	58.6%
83	35.3%	80.9%	71.3%	75.8%	66.7%	36.7%	43.5%	58.6%
82	35.3%	81.1%	71.3%	75.8%	66.6%	36.7%	44.3%	58.7%
81	35.1%	80.9%	71.2%	75.8%	66.4%	36.7%	44.3%	58.6%

threshold	length	ascender	descender	dot	cross	first	last	average
80	35.7%	80.4%	71.2%	75.5%	66.4%	36.7%	44.0%	58.6%
79	35.4%	80.6%	71.5%	75.8%	66.1%	37.1%	44.2%	58.6%
78	35.7%	80.6%	71.5%	76.1%	66.4%	37.2%	44.3%	58.8%
77	36.4%	80.4%	71.8%	76.1%	66.1%	37.2%	44.5%	58.9%
76	36.2%	80.4%	71.8%	75.5%	66.1%	37.1%	45.0%	58.9%
75	36.9%	80.4%	71.7%	75.8%	66.6%	37.2%	45.1%	59.1%
74	37.1%	80.6%	71.8%	76.3%	66.6%	37.6%	45.6%	59.4%
73	37.2%	80.9%	71.7%	76.3%	66.4%	37.7%	45.3%	59.4%
72	36.9%	81.1%	71.7%	75.9%	66.7%	38.1%	45.3%	59.4%
71	36.7%	80.7%	72.0%	75.9%	66.7%	38.2%	45.1%	59.4%
70	37.1%	80.9%	72.0%	75.8%	66.4%	38.6%	45.1%	59.4%
69	36.9%	80.9%	72.0%	76.1%	66.7%	38.6%	45.1%	59.5%
68	37.1%	80.9%	72.0%	75.8%	66.7%	38.4%	45.0%	59.4%
67	36.9%	81.1%	72.5%	75.8%	66.7%	39.0%	45.6%	59.7%
66	37.1%	80.7%	72.7%	75.9%	66.9%	39.4%	45.3%	59.7%
65	36.7%	80.4%	72.7%	75.9%	66.9%	39.5%	45.5%	59.7%
64	36.6%	80.2%	72.8%	76.4%	67.4%	39.4%	45.6%	59.8%
63	36.4%	80.6%	72.3%	76.3%	67.2%	39.5%	45.6%	59.7%
62	36.9%	80.6%	72.3%	76.3%	67.2%	39.2%	46.0%	59.8%
61	36.9%	80.2%	72.0%	76.1%	67.2%	40.2%	46.3%	59.8%
60	36.9%	80.1%	71.8%	76.1%	67.2%	39.7%	46.0%	59.7%
59	37.6%	80.4%	72.0%	76.4%	67.2%	39.9%	45.8%	59.9%
58	37.2%	80.2%	72.2%	76.6%	66.7%	40.4%	45.6%	59.8%
57	37.4%	81.1%	72.2%	76.3%	66.7%	40.5%	45.3%	59.9%
56	37.2%	81.1%	72.3%	76.4%	67.4%	40.4%	45.6%	60.1%
55	36.6%	81.1%	72.2%	76.3%	67.1%	40.7%	45.6%	59.9%
54	37.4%	80.9%	71.8%	76.3%	67.9%	40.7%	46.0%	60.1%
53	37.4%	80.7%	71.8%	76.6%	67.1%	40.5%	45.6%	60.0%
52	37.7%	81.1%	71.7%	77.1%	67.2%	41.0%	45.8%	60.2%
51	37.9%	80.9%	71.8%	77.1%	67.5%	41.0%	46.8%	60.4%
50	38.2%	81.2%	71.5%	77.4%	67.4%	40.9%	46.6%	60.5%
49	37.9%	80.9%	71.3%	78.1%	67.7%	40.5%	46.0%	60.3%
48	37.7%	81.4%	71.7%	77.9%	67.2%	40.5%	46.3%	60.4%

threshold	length	ascender	descender	dot	cross	first	last	average
47	38.1%	81.7%	71.7%	78.1%	67.5%	40.7%	46.5%	60.6%
46	38.2%	81.5%	71.5%	77.8%	67.1%	40.5%	46.5%	60.4%
45	38.4%	81.2%	71.5%	78.1%	67.4%	40.4%	46.1%	60.4%
44	38.4%	81.1%	72.0%	78.6%	68.2%	40.0%	46.1%	60.6%
43	38.2%	81.1%	71.8%	78.6%	68.2%	40.2%	46.1%	60.6%
42	38.7%	81.1%	72.3%	78.6%	67.7%	40.5%	46.1%	60.7%
41	38.6%	80.2%	71.7%	78.4%	68.9%	41.4%	46.3%	60.8%
40	38.9%	80.6%	71.0%	78.7%	68.4%	42.0%	46.3%	60.8%
39	38.6%	80.6%	71.3%	79.2%	68.4%	42.2%	46.6%	61.0%
38	38.7%	80.4%	72.3%	79.6%	68.9%	41.8%	46.8%	61.2%
37	38.7%	80.6%	72.2%	79.7%	69.2%	41.8%	46.3%	61.2%
36	38.4%	80.6%	72.0%	79.7%	68.9%	41.8%	46.3%	61.1%
35	38.9%	80.9%	72.0%	79.6%	69.4%	41.8%	46.5%	61.3%
34	38.7%	80.9%	71.8%	79.6%	69.4%	42.0%	46.3%	61.2%
33	38.4%	80.9%	71.3%	79.4%	69.4%	41.8%	46.3%	61.1%
32	38.6%	80.9%	71.5%	79.7%	69.2%	41.5%	46.0%	61.0%
31	38.6%	81.1%	71.8%	80.1%	69.2%	41.2%	45.6%	61.1%
30	38.7%	81.1%	72.0%	79.9%	69.0%	41.2%	45.6%	61.1%
29	38.6%	81.1%	72.7%	79.7%	69.0%	41.0%	45.8%	61.1%
28	38.9%	80.9%	72.3%	80.1%	69.0%	41.4%	45.8%	61.2%
27	38.6%	80.7%	71.8%	79.9%	68.7%	41.2%	45.8%	61.0%
26	39.0%	80.7%	72.2%	79.9%	68.9%	41.0%	45.8%	61.1%
25	39.4%	80.7%	72.2%	79.9%	68.7%	41.5%	45.8%	61.2%
24	39.2%	80.7%	71.5%	79.9%	68.7%	41.5%	46.1%	61.1%
23	38.7%	80.7%	71.5%	80.1%	68.7%	41.7%	45.8%	61.0%
22	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
21	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
20	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
19	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
18	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
17	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
16	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
15	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%

threshold	length	ascender	descender	dot	cross	first	last	average
14	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
13	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
12	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
11	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
10	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
9	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
8	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
7	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
6	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
5	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
4	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
3	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
2	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
1	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%

Table J-6: Percent correct of the cues: score, with weighting, by threshold: partial data set

threshold	length	ascender	descender	dot	cross	first	last	average
0	33.4%	80.4%	71.8%	74.0%	66.1%	35.1%	41.2%	57.4%
1	33.6%	80.4%	72.3%	74.3%	66.1%	36.1%	41.8%	57.8%
2	34.3%	80.6%	72.5%	75.3%	66.2%	36.2%	42.0%	58.2%
3	34.3%	80.9%	72.0%	75.6%	66.4%	36.7%	42.7%	58.4%
4	34.6%	81.1%	71.7%	75.9%	66.6%	36.4%	43.0%	58.5%
5	35.9%	80.7%	71.7%	77.1%	66.4%	37.1%	43.7%	58.9%
6	36.2%	80.7%	71.5%	77.1%	66.6%	37.1%	43.8%	59.0%
7	37.1%	80.7%	71.0%	77.6%	66.6%	37.6%	44.3%	59.3%
8	37.2%	80.9%	71.0%	77.8%	66.7%	37.6%	44.6%	59.4%
9	37.4%	80.7%	71.0%	77.4%	66.4%	38.2%	45.0%	59.4%
10	37.7%	80.7%	71.2%	78.1%	66.6%	37.9%	45.5%	59.7%
11	37.7%	81.2%	71.3%	78.1%	66.4%	38.2%	45.1%	59.7%
12	37.4%	81.2%	71.2%	78.1%	66.1%	38.6%	45.3%	59.7%
13	37.7%	81.2%	71.3%	78.6%	66.4%	39.7%	45.6%	60.1%
14	37.2%	81.4%	71.7%	78.4%	66.9%	39.5%	45.6%	60.1%
15	38.2%	81.4%	71.7%	78.9%	67.1%	39.9%	45.1%	60.3%
16	38.4%	81.4%	71.8%	79.2%	67.4%	40.4%	45.3%	60.6%
17	38.4%	81.4%	71.8%	79.2%	67.1%	40.7%	45.1%	60.5%
18	37.7%	81.5%	71.8%	79.1%	67.4%	40.9%	46.1%	60.6%
19	37.7%	81.4%	71.8%	79.4%	66.9%	40.5%	46.1%	60.6%
20	38.6%	80.9%	71.5%	79.2%	66.4%	40.9%	46.1%	60.5%

threshold	length	ascender	descender	dot	cross	first	last	average
21	38.1%	81.4%	71.5%	79.6%	65.9%	41.4%	46.1%	60.6%
22	38.4%	81.2%	71.8%	80.1%	66.1%	41.4%	46.6%	60.8%
23	38.9%	81.2%	72.3%	79.9%	65.4%	41.7%	47.0%	60.9%
24	39.0%	80.9%	72.7%	79.1%	65.6%	41.8%	47.1%	60.9%
25	39.4%	80.9%	72.7%	79.4%	66.2%	41.7%	47.6%	61.1%
26	38.4%	81.1%	72.3%	79.6%	66.9%	42.0%	47.8%	61.1%
27	39.2%	81.2%	71.8%	79.4%	66.4%	42.0%	47.1%	61.0%
28	38.9%	81.4%	71.3%	79.1%	67.1%	42.3%	47.0%	61.0%
29	38.6%	81.2%	71.7%	78.7%	67.5%	42.3%	47.1%	61.0%
30	39.2%	81.5%	71.5%	78.7%	66.9%	42.5%	46.5%	61.0%
31	38.7%	81.9%	71.8%	78.4%	66.7%	42.3%	46.5%	60.9%
32	38.6%	81.9%	71.2%	78.4%	67.5%	42.2%	46.3%	60.9%
33	39.0%	82.4%	71.5%	78.4%	67.7%	42.7%	46.5%	61.2%
34	38.7%	82.4%	71.5%	79.1%	67.9%	43.0%	46.5%	61.3%
35	38.4%	81.9%	71.5%	78.9%	68.0%	42.8%	46.6%	61.2%
36	38.4%	81.9%	71.7%	79.1%	68.7%	42.7%	46.8%	61.3%
37	38.7%	82.2%	71.2%	78.7%	68.5%	42.3%	46.8%	61.2%
38	38.9%	82.2%	71.3%	78.7%	69.0%	42.5%	47.4%	61.4%
39	38.9%	82.0%	71.0%	78.7%	69.0%	43.2%	47.4%	61.5%
40	39.4%	81.9%	71.3%	78.9%	69.0%	42.7%	47.4%	61.5%
41	39.4%	82.0%	72.2%	78.9%	68.5%	42.7%	47.1%	61.5%
42	39.5%	82.0%	72.2%	79.4%	68.4%	42.7%	47.0%	61.6%
43	38.6%	82.7%	72.0%	79.2%	68.0%	43.0%	47.3%	61.5%
44	38.9%	82.5%	72.3%	79.4%	68.5%	42.5%	47.4%	61.7%
45	38.6%	82.5%	71.8%	79.2%	68.5%	42.7%	47.3%	61.5%
46	39.2%	82.5%	72.0%	79.2%	69.2%	42.7%	47.3%	61.7%
47	38.9%	82.4%	71.5%	79.2%	68.4%	42.5%	47.0%	61.4%
48	38.6%	82.5%	71.5%	79.4%	67.7%	42.3%	47.1%	61.3%
49	38.6%	82.4%	71.3%	79.4%	68.0%	42.2%	47.9%	61.4%
50	38.6%	82.4%	71.3%	79.6%	67.7%	42.2%	47.8%	61.4%
51	38.4%	81.9%	71.2%	79.9%	67.9%	41.7%	47.3%	61.2%
52	37.9%	82.4%	70.8%	79.9%	67.7%	41.5%	47.0%	61.0%
53	37.9%	82.5%	70.8%	80.1%	68.2%	41.7%	47.3%	61.2%

threshold	length	ascender	descender	dot	cross	first	last	average
54	38.1%	82.4%	70.5%	79.9%	68.2%	41.5%	47.1%	61.1%
55	38.2%	81.9%	70.8%	79.9%	68.2%	41.5%	47.1%	61.1%
56	38.2%	81.7%	71.3%	80.1%	68.4%	41.4%	47.1%	61.2%
57	38.7%	81.7%	71.5%	79.9%	68.4%	41.4%	47.3%	61.3%
58	38.7%	81.5%	71.5%	79.9%	68.2%	41.2%	47.0%	61.1%
59	38.6%	80.9%	71.3%	79.9%	68.4%	41.5%	46.6%	61.0%
60	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
61	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
62	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
63	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
64	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
65	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
66	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
67	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
68	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
69	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
70	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
71	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
72	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
73	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
74	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
75	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
76	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
77	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
78	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
79	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
80	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
81	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
82	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
83	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
84	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
85	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
86	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%

threshold	length	ascender	descender	dot	cross	first	last	average
87	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
88	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
89	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
90	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
91	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
92	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
93	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
94	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
95	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
96	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
97	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
98	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
99	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%

Table J-7: Percent correct of the cues: difference, with weighting, by threshold: partial data set

threshold	length	ascender	descender	dot	cross	first	last	average
99	33.4%	80.4%	71.8%	74.0%	66.1%	35.1%	41.2%	57.4%
98	33.6%	80.4%	72.2%	74.3%	66.1%	35.6%	41.8%	57.7%
97	33.9%	80.6%	72.0%	74.6%	66.1%	35.7%	42.0%	57.8%
96	34.1%	80.6%	72.3%	74.8%	66.1%	36.1%	42.5%	58.1%
95	33.9%	80.9%	72.0%	75.3%	66.1%	36.1%	42.5%	58.1%
94	35.1%	80.9%	72.0%	75.6%	65.9%	36.1%	43.2%	58.4%
93	35.4%	80.7%	71.7%	76.3%	66.2%	36.4%	43.2%	58.6%
92	35.9%	80.7%	71.8%	76.8%	66.2%	36.7%	43.3%	58.8%
91	36.1%	81.1%	72.0%	77.1%	66.6%	37.2%	44.0%	59.1%
90	36.7%	80.9%	71.8%	77.1%	66.4%	37.6%	44.2%	59.2%
89	37.1%	80.7%	71.5%	77.4%	66.6%	37.1%	45.1%	59.4%
88	37.1%	80.7%	71.2%	77.6%	66.9%	37.2%	45.3%	59.4%
87	36.9%	80.7%	71.2%	77.8%	66.7%	37.6%	45.1%	59.4%
86	36.7%	80.9%	71.3%	77.8%	66.7%	38.4%	44.8%	59.5%
85	36.7%	81.1%	70.8%	77.4%	66.9%	38.4%	44.8%	59.4%
84	37.9%	81.1%	71.0%	77.6%	67.2%	38.7%	44.8%	59.8%
83	37.9%	80.9%	71.3%	77.6%	67.9%	38.6%	45.1%	59.9%
82	37.2%	81.1%	71.5%	77.6%	68.0%	39.2%	45.3%	60.0%
81	37.1%	81.2%	71.8%	78.1%	67.7%	39.4%	46.0%	60.2%
80	36.9%	81.1%	71.5%	78.3%	67.4%	39.5%	45.8%	60.1%
79	37.7%	80.6%	71.7%	78.4%	67.2%	39.7%	45.5%	60.1%

threshold	length	ascender	descender	dot	cross	first	last	average
78	37.9%	80.9%	71.8%	78.4%	66.7%	40.2%	46.1%	60.3%
77	38.2%	80.7%	72.0%	78.9%	67.2%	40.7%	46.1%	60.6%
76	38.9%	80.7%	73.0%	79.2%	67.4%	41.0%	46.0%	60.9%
75	38.7%	80.7%	73.1%	78.7%	67.1%	41.2%	46.3%	60.8%
74	39.2%	80.6%	72.8%	79.2%	67.9%	41.5%	46.3%	61.1%
73	39.5%	80.9%	72.8%	79.6%	67.5%	42.0%	46.5%	61.3%
72	40.0%	81.1%	72.2%	79.4%	67.4%	42.2%	46.3%	61.2%
71	39.7%	80.9%	72.0%	79.1%	67.4%	41.8%	46.1%	61.0%
70	39.7%	80.9%	72.5%	79.1%	67.5%	42.0%	46.3%	61.1%
69	40.0%	81.2%	72.8%	79.2%	66.9%	42.5%	45.8%	61.2%
68	39.9%	81.2%	72.8%	79.1%	67.1%	42.0%	46.0%	61.1%
67	40.0%	81.4%	72.7%	78.7%	66.7%	42.0%	45.6%	61.0%
66	39.9%	81.9%	72.5%	78.6%	66.2%	42.5%	46.8%	61.2%
65	39.4%	81.7%	72.7%	78.7%	66.9%	43.0%	47.0%	61.3%
64	39.0%	81.5%	72.2%	78.7%	67.1%	42.8%	47.4%	61.3%
63	38.9%	81.5%	72.2%	78.4%	67.4%	42.7%	47.8%	61.3%
62	38.4%	82.0%	72.0%	78.4%	67.4%	42.8%	47.4%	61.2%
61	38.7%	82.0%	71.8%	78.6%	67.7%	42.5%	47.6%	61.3%
60	38.9%	81.9%	71.3%	78.7%	67.4%	43.3%	47.4%	61.3%
59	38.9%	81.5%	71.3%	79.1%	67.4%	43.0%	47.3%	61.2%
58	39.4%	81.7%	71.3%	79.2%	67.4%	42.3%	47.1%	61.2%
57	39.2%	81.7%	71.3%	79.4%	67.7%	42.2%	47.4%	61.3%
56	38.7%	82.2%	71.0%	79.1%	67.7%	42.5%	47.8%	61.3%
55	39.0%	82.0%	70.8%	79.4%	67.5%	42.2%	47.8%	61.3%
54	38.2%	82.0%	70.8%	79.2%	67.5%	42.5%	47.3%	61.1%
53	38.9%	81.9%	71.0%	79.2%	68.4%	42.8%	47.4%	61.4%
52	38.6%	82.0%	71.3%	79.4%	67.5%	43.0%	47.1%	61.3%
51	38.6%	82.2%	71.5%	79.4%	67.7%	42.8%	47.0%	61.3%
50	38.4%	82.0%	71.7%	79.6%	67.7%	43.0%	47.4%	61.4%
49	38.9%	82.2%	71.5%	79.9%	67.2%	42.5%	47.3%	61.4%
48	38.2%	81.7%	71.7%	80.1%	67.5%	42.0%	46.5%	61.1%
47	38.2%	82.2%	71.2%	79.9%	67.7%	41.8%	46.6%	61.1%
46	38.6%	82.4%	71.2%	80.2%	68.0%	42.0%	46.8%	61.3%

threshold	length	ascender	descender	dot	cross	first	last	average
45	38.4%	82.4%	70.8%	79.9%	67.9%	41.8%	46.6%	61.1%
44	39.0%	81.9%	71.0%	79.9%	68.0%	41.8%	46.6%	61.2%
43	39.0%	81.7%	71.7%	80.1%	68.4%	41.8%	47.1%	61.4%
42	39.0%	81.7%	71.8%	79.9%	68.4%	41.8%	47.3%	61.4%
41	38.9%	81.5%	72.2%	79.9%	68.2%	41.5%	46.8%	61.3%
40	38.7%	80.9%	72.0%	79.9%	68.5%	41.8%	46.6%	61.2%
39	38.9%	81.1%	71.7%	80.1%	68.5%	42.0%	46.5%	61.2%
38	38.6%	81.1%	71.8%	80.1%	68.9%	41.8%	46.5%	61.2%
37	38.6%	81.1%	71.7%	80.1%	68.9%	41.8%	46.5%	61.2%
36	38.4%	81.1%	71.7%	80.1%	68.7%	41.8%	46.5%	61.2%
35	38.4%	80.9%	71.7%	79.9%	68.7%	42.0%	46.5%	61.1%
34	38.7%	80.9%	71.7%	79.9%	68.4%	42.0%	46.3%	61.1%
33	38.7%	80.9%	71.7%	79.9%	68.5%	41.8%	46.3%	61.1%
32	38.7%	80.9%	71.7%	79.9%	68.7%	41.8%	46.3%	61.1%
31	38.7%	80.9%	71.5%	79.9%	68.7%	41.8%	46.3%	61.1%
30	38.7%	80.9%	71.5%	79.9%	68.7%	41.8%	46.5%	61.1%
29	38.7%	80.9%	71.5%	80.1%	68.7%	41.8%	46.3%	61.1%
28	38.7%	80.9%	71.5%	80.1%	68.7%	41.8%	46.3%	61.1%
27	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
26	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
25	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
24	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
23	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
22	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
21	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
20	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
19	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
18	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
17	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
16	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
15	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
14	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
13	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%

threshold	length	ascender	descender	dot	cross	first	last	average
12	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
11	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
10	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
9	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
8	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
7	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
6	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
5	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
4	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
3	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
2	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
1	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%

Table J-8: Percent correct of the cues: ratio, with weighting, by threshold: partial data set

threshold	length	ascender	descender	dot	cross	first	last	average
0.0	36.9%	80.1%	70.5%	80.4%	68.4%	40.2%	45.5%	60.3%
0.1	38.4%	80.9%	70.2%	80.2%	68.5%	40.5%	45.3%	60.6%
0.2	37.6%	80.1%	70.5%	79.7%	68.4%	41.2%	46.3%	60.5%
0.3	37.4%	80.4%	71.2%	80.6%	68.0%	40.5%	46.5%	60.6%
0.4	37.7%	80.6%	70.8%	80.1%	68.2%	40.9%	46.1%	60.6%
0.5	38.1%	80.6%	70.5%	80.2%	68.4%	40.7%	46.8%	60.7%
0.6	37.6%	80.2%	70.7%	80.1%	68.2%	41.0%	46.6%	60.6%
0.7	38.1%	80.2%	70.5%	79.7%	68.2%	41.2%	46.5%	60.6%
0.8	38.2%	80.6%	70.7%	79.9%	68.5%	41.5%	46.8%	60.9%
0.9	38.6%	80.7%	71.0%	79.6%	68.7%	41.8%	46.5%	61.0%
1.0	38.6%	80.9%	71.3%	80.1%	68.7%	41.8%	46.5%	61.1%
1.1	38.4%	81.1%	71.5%	79.9%	68.7%	42.0%	46.6%	61.2%
1.2	38.9%	81.1%	71.5%	80.1%	68.5%	42.0%	47.0%	61.3%
1.3	38.9%	81.2%	71.5%	80.2%	68.5%	42.0%	47.0%	61.3%
1.4	38.2%	81.5%	71.7%	80.4%	69.0%	42.2%	47.0%	61.4%
1.5	38.6%	81.7%	72.0%	80.6%	68.9%	42.2%	47.0%	61.5%
1.6	38.7%	81.9%	72.0%	80.6%	68.9%	41.7%	47.0%	61.5%
1.7	39.2%	82.0%	72.3%	80.4%	69.2%	41.7%	47.4%	61.8%
1.8	39.4%	82.2%	72.3%	80.4%	69.0%	41.8%	47.6%	61.8%
1.9	39.5%	82.4%	72.3%	80.2%	68.4%	41.8%	47.9%	61.8%
2.0	39.5%	82.4%	72.2%	80.2%	68.7%	41.8%	47.9%	61.8%
2.1	39.5%	82.4%	72.3%	80.4%	68.7%	42.3%	47.8%	61.9%
2.2	39.7%	82.4%	72.2%	80.4%	68.5%	42.3%	48.4%	62.0%
2.3	39.2%	82.5%	72.3%	80.2%	68.2%	42.7%	48.3%	61.9%
2.4	39.2%	82.5%	72.5%	80.2%	67.9%	42.7%	48.4%	61.9%
2.5	38.7%	82.4%	72.7%	80.1%	67.9%	42.7%	48.6%	61.8%
2.6	38.7%	82.2%	72.5%	80.1%	67.7%	42.3%	48.6%	61.7%
2.7	38.7%	82.2%	72.8%	79.9%	67.9%	42.5%	48.8%	61.8%
2.8	38.9%	82.5%	73.3%	79.9%	68.2%	42.5%	48.6%	62.0%

threshold	length	ascender	descender	dot	cross	first	last	average
2.9	39.0%	82.7%	73.3%	79.7%	68.4%	42.5%	48.6%	62.0%
3.0	38.7%	82.9%	73.3%	79.7%	68.2%	42.7%	48.6%	62.0%
3.1	38.7%	82.7%	73.3%	79.7%	67.9%	42.5%	48.4%	61.9%
3.2	39.0%	83.0%	73.1%	79.7%	67.9%	42.3%	48.4%	61.9%
3.3	38.9%	83.0%	73.0%	79.7%	67.9%	42.5%	48.6%	61.9%
3.4	36.9%	80.1%	70.5%	80.4%	68.4%	40.2%	45.5%	60.3%
3.5	38.4%	80.9%	70.0%	80.2%	68.5%	40.4%	45.5%	60.6%

Table J-9: Percent correct of the cues: power, by threshold: partial data set

threshold	length	ascender	descender	dot	cross	first	last	average
0.0	36.9%	80.1%	70.5%	80.4%	68.4%	40.2%	45.5%	60.3%
0.1	37.9%	80.7%	70.3%	80.6%	69.0%	41.0%	45.6%	60.7%
0.2	36.9%	80.1%	70.2%	80.2%	68.4%	40.9%	45.8%	60.3%
0.3	37.4%	80.2%	70.3%	80.2%	68.2%	40.2%	46.5%	60.4%
0.4	37.7%	80.4%	70.2%	80.4%	67.9%	40.5%	46.0%	60.4%
0.5	37.7%	80.4%	70.5%	80.4%	68.5%	40.9%	46.5%	60.7%
0.6	38.1%	80.2%	70.7%	80.1%	68.2%	41.0%	46.6%	60.7%
0.7	38.1%	80.4%	70.5%	79.7%	68.2%	41.4%	46.6%	60.7%
0.8	38.2%	80.6%	70.7%	79.9%	68.5%	41.4%	46.6%	60.8%
0.9	38.4%	80.7%	70.8%	79.7%	68.7%	41.8%	46.8%	61.0%
1.0	38.6%	80.9%	71.5%	80.1%	68.7%	41.8%	46.5%	61.1%
1.1	38.4%	81.1%	71.5%	79.9%	68.7%	42.0%	46.6%	61.2%
1.2	38.9%	81.1%	71.5%	80.1%	68.5%	42.0%	47.0%	61.3%
1.3	38.9%	81.2%	71.3%	80.2%	68.7%	42.0%	47.0%	61.3%
1.4	38.4%	81.5%	71.7%	80.4%	69.0%	42.2%	47.0%	61.4%
1.5	38.6%	81.7%	72.0%	80.6%	68.7%	42.2%	47.0%	61.5%
1.6	38.7%	81.9%	72.0%	80.6%	68.9%	41.8%	47.0%	61.5%
1.7	39.2%	82.0%	72.2%	80.4%	69.2%	41.7%	47.4%	61.7%
1.8	39.4%	82.0%	72.3%	80.2%	69.0%	41.8%	47.6%	61.8%
1.9	39.5%	82.4%	72.5%	80.2%	68.4%	41.8%	47.8%	61.8%
2.0	39.7%	82.4%	72.2%	80.2%	68.5%	42.0%	47.9%	61.8%
2.1	39.5%	82.4%	72.2%	80.4%	68.7%	42.3%	47.9%	61.9%
2.2	39.5%	82.4%	72.2%	80.4%	68.7%	42.3%	48.3%	62.0%
2.3	39.2%	82.5%	72.3%	80.2%	68.2%	42.7%	48.3%	61.9%
2.4	39.2%	82.5%	72.5%	80.2%	68.0%	42.7%	48.4%	61.9%
2.5	38.7%	82.4%	72.7%	80.1%	67.9%	42.7%	48.6%	61.8%
2.6	38.7%	82.2%	72.5%	79.9%	67.7%	42.5%	48.6%	61.7%
2.7	38.7%	82.2%	72.7%	79.9%	67.9%	42.5%	48.8%	61.8%
2.8	38.9%	82.5%	73.3%	79.9%	68.2%	42.5%	48.6%	62.0%
2.9	39.0%	82.7%	73.3%	79.7%	68.4%	42.5%	48.6%	62.0%
3.0	38.6%	82.9%	73.3%	79.7%	68.2%	42.7%	48.6%	62.0%

threshold	length	ascender	descender	dot	cross	first	last	average
3.1	38.9%	82.7%	73.3%	79.7%	67.9%	42.3%	48.4%	61.9%
3.2	38.9%	83.0%	73.1%	79.7%	68.0%	42.3%	48.4%	61.9%
3.3	38.7%	83.0%	73.0%	79.7%	67.7%	42.5%	48.6%	61.9%
3.4	36.9%	80.1%	70.5%	80.4%	68.4%	40.2%	45.5%	60.3%
3.5	38.1%	80.7%	70.3%	80.6%	69.0%	41.0%	45.6%	60.8%

Table J-10: Percent correct of the cues: power + normalization, by threshold: partial data set

The final ten tables Table J-11 to Table J-20 show results only for the chosen threshold used to limit the number of candidates or to adjust the confidence score of the candidates. However, results for all of the other combination of methods are provided. Results are given by the combination of methods used to calculate the results. A dash means that the combination of methods cannot be used for that cue. Table J-11 to Table J-14 present results for rank, score, difference and ratio when weighting was not used. Table J-15 to Table J-18 present results for rank, score, difference and ratio when weighting was used. Table J-19 and Table J-20 show results for power and power plus normalization.

Mean, median and mode were the three ways used to calculate values (see section 4.3.2).

Average, reduce and initial were the three ways used to force a single-valued outcome (see section 4.3.3).

Bias is the use of bias correction (see section 5.2.2).

Associations is the use of word associations (see section 5.2.5).

method	length	ascender	descender	dot	cross	first	last
mean	37.2%	80.2%	69.2%	75.6%	67.2%	-	-
mean, bias	-	-	71.8%	79.7%	-	-	-
median, average	36.4%	80.2%	69.2%	75.6%	67.2%	-	-
median, average, bias	-	-	71.7%	75.6%	-	-	-
median, reduce	36.6%	80.7%	70.5%	73.1%	67.4%	-	-
median, initial	34.8%	80.9%	72.0%	73.6%	68.0%	-	-
mode, average	36.2%	80.2%	69.2%	75.6%	67.2%	-	-
mode, average, bias	-	-	71.7%	75.6%	-	-	-
mode, reduce	35.4%	80.7%	70.5%	73.1%	67.4%	37.6%	42.0%
mode, initial	35.9%	80.9%	72.0%	73.6%	68.0%	38.4%	41.0%

Table J-11: Percent correct of the cues: rank, without weighting, by combination of methods: partial data set

method	length	ascender	descender	dot	cross	first	last
mean	37.2%	80.2%	69.2%	75.6%	67.2%	-	-
mean, bias	-	-	71.8%	79.7%	-	-	-
median, average	36.4%	80.2%	69.2%	75.6%	67.2%	-	-
median, average, bias	-	-	71.7%	75.6%	-	-	-
median, reduce	36.6%	80.7%	70.5%	73.1%	67.4%	-	-
median, initial	34.8%	80.9%	72.0%	73.6%	68.0%	-	-
mode, average	36.2%	80.2%	69.2%	75.6%	67.2%	-	-
mode, average, bias	-	-	71.7%	75.6%	-	-	-
mode, reduce	35.4%	80.7%	70.5%	73.1%	67.4%	37.6%	42.0%
mode, initial	35.9%	80.9%	72.0%	73.6%	68.0%	38.4%	41.0%

Table J-12: Percent correct of the cues: score, without weighting, by combination of methods: partial data set

method	length	ascender	descender	dot	cross	first	last
mean	37.2%	80.2%	69.2%	75.6%	67.2%	-	-
mean, bias	-	-	71.8%	79.7%	-	-	-
median, average	36.4%	80.2%	69.2%	75.6%	67.2%	-	-
median, average, bias	-	-	71.7%	75.6%	-	-	-
median, reduce	36.6%	80.7%	70.5%	73.1%	67.4%	-	-
median, initial	34.8%	80.9%	72.0%	73.6%	68.0%	-	-
mode, average	36.2%	80.2%	69.2%	75.6%	67.2%	-	-
mode, average, bias	-	-	71.7%	75.6%	-	-	-
mode, reduce	35.4%	80.7%	70.5%	73.1%	67.4%	37.6%	42.0%
mode, initial	35.9%	80.9%	72.0%	73.6%	68.0%	38.4%	41.0%

Table J-13: Percent correct of the cues: difference, without weighting, by combination of methods: partial data set

method	length	ascender	descender	dot	cross	first	last
mean	39.2%	80.4%	69.5%	76.1%	67.5%	-	-
mean, bias	-	-	72.7%	80.1%	-	-	-
median, average	37.7%	80.4%	69.5%	76.1%	67.5%	-	-
median, average, bias	-	-	72.5%	76.1%	-	-	-
median, reduce	38.6%	80.7%	72.0%	74.1%	67.2%	-	-
median, initial	37.2%	80.7%	71.8%	73.8%	67.5%	-	-
mode, average	37.7%	80.4%	69.5%	76.1%	67.5%	-	-
mode, average, bias	-	-	72.5%	76.1%	-	-	-
mode, reduce	38.1%	80.7%	72.0%	74.1%	67.2%	36.9%	40.7%
mode, initial	37.6%	80.7%	71.8%	73.8%	67.5%	37.6%	41.4%

Table J-14: Percent correct of the cues: ratio, without weighting, by combination of methods: partial data set

method	length	ascender	descender	dot	cross	first	last
mean	37.4%	81.5%	71.2%	73.5%	69.0%	-	-
mean, bias	-	-	72.3%	80.2%	-	-	-
median, average	38.4%	81.5%	71.2%	73.5%	69.0%	-	-
median, average, bias	-	-	71.7%	73.5%	-	-	-
median, reduce	38.4%	81.5%	71.5%	73.5%	68.9%	-	-
median, initial	38.4%	81.5%	71.7%	73.5%	68.9%	-	-
mode, average	37.4%	81.5%	71.2%	73.5%	69.0%	-	-
mode, average, bias	-	-	71.7%	73.5%	-	-	-
mode, reduce	37.2%	81.5%	71.5%	73.5%	68.9%	36.1%	42.0%
mode, reduce, associations	-	-	-	-	-	42.7%	47.6%
mode, initial	37.2%	81.5%	71.7%	73.5%	68.9%	36.1%	42.0%
mode, initial, associations	-	-	-	-	-	42.7%	47.6%

Table J-15: Percent correct of the cues: rank, with weighting, by combination of methods: partial data set

method	length	ascender	descender	dot	cross	first	last
mean	38.4%	80.9%	69.4%	74.5%	69.4%	-	-
mean, bias	-	80.9%	72.0%	79.6%	69.4%	-	-
median, average	38.9%	80.9%	69.4%	74.5%	69.4%	-	-
median, average, bias	-	80.9%	70.7%	74.5%	69.4%	-	-
median, reduce	38.9%	80.9%	70.0%	74.5%	69.2%	-	-
median, initial	38.9%	80.9%	70.0%	74.5%	69.2%	-	-
mode, average	36.4%	80.9%	69.4%	74.5%	69.4%	-	-
mode, average, bias	-	80.9%	70.7%	74.5%	69.4%	-	-
mode, reduce	36.4%	80.9%	70.0%	74.5%	69.2%	36.6%	43.0%
mode, reduce, associations	-	-	-	-	-	41.8%	46.5%
mode, initial	36.4%	80.9%	70.0%	74.5%	69.2%	36.6%	43.0%
mode, initial, associations	-	-	-	-	-	41.7%	46.5%

Table J-16: Percent correct of the cues: score, with weighting, by combination of methods: partial data set

method	length	ascender	descender	dot	cross	first	last
mean	37.2%	82.5%	70.7%	73.5%	68.5%	-	-
mean, bias	-	82.5%	72.3%	79.4%	68.5%	-	-
median, average	38.9%	82.5%	70.7%	73.5%	68.5%	-	-
median, average, bias	-	82.5%	71.2%	73.5%	68.5%	-	-
median, reduce	38.7%	82.5%	71.0%	73.5%	68.4%	-	-
median, initial	38.9%	82.5%	71.2%	73.5%	68.4%	-	-
mode, average	38.1%	82.5%	70.7%	73.5%	68.5%	-	-
mode, average, bias	-	82.5%	71.2%	73.5%	68.5%	-	-
mode, reduce	37.9%	82.5%	71.0%	73.5%	68.4%	37.1%	41.7%
mode, reduce, associations	-	-	-	-	-	42.5%	47.4%
mode, initial	37.7%	82.5%	71.2%	73.5%	68.4%	37.1%	41.7%
mode, initial, associations	-	-	-	-	-	42.5%	47.4%

Table J-17: Percent correct of the cues: difference, with weighting, by combination of methods: partial data set

method	length	ascender	descender	dot	cross	first	last
mean	37.9%	81.9%	70.3%	73.5%	68.4%	-	-
mean, bias	-	81.9%	71.0%	79.2%	68.4%	-	-
median, average	38.9%	81.9%	70.3%	73.5%	68.4%	-	-
median, average, bias	-	81.9%	70.7%	73.5%	68.4%	-	-
median, reduce	38.9%	81.9%	70.7%	73.5%	68.2%	-	-
median, initial	38.9%	81.9%	70.7%	73.5%	68.2%	-	-
mode, average	37.7%	81.9%	70.3%	73.5%	68.4%	-	-
mode, average, bias	-	81.9%	70.7%	73.5%	68.4%	-	-
mode, reduce	37.6%	81.9%	70.7%	73.5%	68.2%	37.2%	42.2%
mode, reduce, associations	-	-	-	-	-	42.8%	47.4%
mode, initial	37.4%	81.9%	70.7%	73.5%	68.2%	37.4%	42.3%
mode, initial, associations	-	-	-	-	-	42.8%	47.4%

Table J-18: Percent correct of the cues: ratio, with weighting, by combination of methods: partial data set

method	length	ascender	descender	dot	cross	first	last
mean	37.9%	82.7%	70.7%	75.0%	68.4%	-	-
mean, bias	-	82.7%	73.3%	79.7%	68.4%	-	-
median, average	39.0%	82.7%	70.7%	75.0%	68.4%	-	-
median, average, bias	-	82.7%	71.0%	75.0%	68.4%	-	-
median, reduce	39.0%	82.7%	71.0%	75.0%	68.2%	-	-
median, initial	39.0%	82.7%	71.0%	75.0%	68.2%	-	-
mode, average	38.6%	82.7%	70.7%	75.0%	68.4%	-	-
mode, average, bias	-	82.7%	71.0%	75.0%	68.4%	-	-
mode, reduce	38.4%	82.7%	71.0%	75.0%	68.2%	37.6%	41.8%
mode, reduce, associations	-	-	-	-	-	42.5%	48.6%
mode, initial	38.4%	82.7%	71.0%	75.0%	68.2%	37.6%	41.8%
mode, initial, associations	-	-	-	-	-	42.5%	48.6%

Table J-19: Percent correct of the cues: power, by combination of methods: partial data set

method	length	ascender	descender	dot	cross	first	last
mean	38.7%	82.4%	70.0%	74.6%	68.7%	-	-
mean, bias	-	82.4%	72.2%	80.4%	68.7%	-	-
median, average	39.5%	82.4%	70.0%	74.6%	68.7%	-	-
median, average, bias	-	82.4%	70.3%	74.6%	68.7%	-	-
median, reduce	39.5%	82.4%	70.3%	74.6%	68.5%	-	-
median, initial	39.5%	82.4%	70.3%	74.6%	68.5%	-	-
mode, average	37.7%	82.4%	70.0%	74.6%	68.7%	-	-
mode, average, bias	-	82.4%	70.3%	74.6%	68.7%	-	-
mode, reduce	37.6%	82.4%	70.3%	74.6%	68.5%	37.6%	41.8%
mode, reduce, associations	-	-	-	-	-	42.3%	48.3%
mode, initial	37.6%	82.4%	70.3%	74.6%	68.5%	37.6%	41.8%
mode, initial, associations	-	-	-	-	-	42.3%	48.3%

Table J-20: Percent correct of the cues: power + normalization, by combination of methods: partial data set

Appendix K: Merging Direct Cue Extraction and using the Candidate List

Table K-1 shows the effect on the recognition rate of a range of multipliers applied to the probability scores (see section 5.4.2).

increase/decrease	ascender	descender	average
0.0	87.8%	93.4%	90.6%
0.1	87.8%	93.2%	90.5%
0.2	87.8%	93.2%	90.5%
0.3	87.8%	93.2%	90.5%
0.4	87.8%	93.1%	90.4%
0.5	88.0%	93.2%	90.6%
0.6	87.8%	93.2%	90.5%
0.7	87.6%	93.2%	90.4%
0.8	87.6%	93.2%	90.4%
0.9	87.6%	93.1%	90.4%
1.0	88.0%	93.4%	90.7%
1.1	87.8%	93.2%	90.5%
1.2	87.8%	93.4%	90.6%
1.3	87.8%	93.4%	90.6%
1.4	87.8%	93.4%	90.6%
1.5	87.8%	93.2%	90.5%
1.6	87.8%	93.2%	90.5%
1.7	88.0%	93.4%	90.7%
1.8	88.0%	93.6%	90.8%
1.9	88.0%	93.6%	90.8%
2.0	88.0%	93.4%	90.7%

increase/decrease	ascender	descender	average
2.1	88.1%	93.6%	90.9%
2.2	88.1%	93.6%	90.9%
2.3	88.1%	93.6%	90.9%
2.4	88.1%	93.6%	90.9%
2.5	88.1%	93.6%	90.9%
2.6	88.3%	93.7%	91.0%
2.7	88.3%	93.7%	91.0%
2.8	88.1%	93.7%	90.9%
2.9	88.3%	93.9%	91.1%
3.0	88.3%	93.9%	91.1%
3.1	88.3%	93.7%	91.0%
3.2	88.3%	93.9%	91.1%
3.3	88.3%	93.9%	91.1%
3.4	88.1%	93.9%	91.0%
3.5	88.1%	93.9%	91.0%
3.6	88.1%	93.9%	91.0%
3.7	88.1%	93.9%	91.0%
3.8	88.1%	94.1%	91.1%
3.9	88.3%	93.9%	91.1%
4.0	88.3%	93.9%	91.1%
4.1	88.3%	93.9%	91.1%
4.2	88.1%	93.7%	90.9%
4.3	88.1%	93.7%	90.9%
4.4	88.0%	93.9%	90.9%
4.5	88.1%	93.9%	91.0%
4.6	88.0%	93.9%	90.9%
4.7	88.0%	93.7%	90.9%
4.8	88.1%	93.7%	90.9%
4.9	88.3%	93.7%	91.0%
5.0	88.5%	93.9%	91.2%
5.1	88.5%	93.9%	91.2%
5.2	88.5%	93.9%	91.2%
5.3	88.5%	93.9%	91.2%

increase/decrease	ascender	descender	average
5.4	88.5%	93.9%	91.2%
5.5	88.5%	93.9%	91.2%
5.6	88.5%	93.9%	91.2%
5.7	88.5%	93.9%	91.2%
5.8	88.5%	93.9%	91.2%
5.9	88.3%	93.9%	91.1%
6.0	88.3%	93.9%	91.1%
6.1	88.3%	93.9%	91.1%
6.2	88.3%	93.9%	91.1%
6.3	88.5%	94.1%	91.3%
6.4	88.5%	93.9%	91.2%
6.5	88.5%	94.1%	91.3%
6.6	88.5%	94.1%	91.3%
6.7	88.5%	94.1%	91.3%
6.8	88.5%	93.9%	91.2%
6.9	88.5%	94.2%	91.4%
7.0	88.6%	94.2%	91.4%
7.1	88.6%	94.2%	91.4%
7.2	88.6%	94.2%	91.4%
7.3	88.6%	94.2%	91.4%
7.4	88.6%	94.2%	91.4%
7.5	88.6%	94.2%	91.4%
7.6	88.6%	94.2%	91.4%
7.7	88.6%	94.1%	91.4%
7.8	88.8%	94.2%	91.5%
7.9	88.8%	94.2%	91.5%
8.0	88.8%	94.2%	91.5%
8.1	88.8%	94.2%	91.5%
8.2	88.8%	94.2%	91.5%
8.3	88.8%	94.2%	91.5%
8.4	88.8%	94.2%	91.5%
8.5	88.8%	94.1%	91.4%
8.6	88.8%	94.1%	91.4%

increase/decrease	ascender	descender	average
8.7	88.8%	94.1%	91.4%
8.8	88.8%	93.9%	91.4%
8.9	89.0%	94.1%	91.5%
9.0	89.0%	94.1%	91.5%
9.1	89.0%	94.1%	91.5%
9.2	89.0%	93.9%	91.4%
9.3	89.1%	94.1%	91.6%
9.4	89.3%	94.1%	91.7%
9.5	89.3%	94.1%	91.7%
9.6	89.3%	93.9%	91.6%
9.7	89.5%	94.1%	91.8%
9.8	89.5%	94.1%	91.8%
9.9	89.5%	94.1%	91.8%
10.0	89.5%	94.1%	91.8%
10.1	89.5%	94.1%	91.8%
10.2	89.3%	94.2%	91.8%
10.3	89.5%	94.2%	91.8%
10.4	89.5%	94.2%	91.8%
10.5	89.5%	94.2%	91.8%
10.6	89.5%	94.2%	91.8%
10.7	89.5%	94.1%	91.8%
10.8	89.6%	94.1%	91.8%
10.9	89.5%	94.2%	91.8%
11.0	89.5%	94.2%	91.8%
11.1	89.5%	94.2%	91.8%
11.2	89.3%	94.1%	91.7%
11.3	89.5%	93.9%	91.7%
11.4	89.5%	93.9%	91.7%
11.5	89.6%	93.9%	91.8%
11.6	89.6%	93.9%	91.8%
11.7	89.6%	93.9%	91.8%
11.8	89.6%	93.7%	91.7%
11.9	89.5%	93.9%	91.7%

increase/decrease	ascender	descender	average
12.0	89.6%	93.9%	91.8%
12.1	89.6%	93.9%	91.8%
12.2	89.6%	93.9%	91.8%
12.3	89.6%	93.9%	91.8%
12.4	89.6%	93.9%	91.8%
12.5	89.6%	93.9%	91.8%
12.6	89.6%	93.7%	91.7%
12.7	89.3%	93.7%	91.5%
12.8	89.3%	93.7%	91.5%
12.9	89.3%	93.4%	91.4%

Table K-1: Merging direct cue extraction and using the candidate list. Percent correct of the cues ascender and descender, results given by the multiplier applied to the probability scores: partial data set

Appendix L: The Effect of Word Frequency

The score given to each candidate by the word level method was multiplied by a value calculated from candidate's word frequency to give a final score (see section 6.3). Different modifications to word frequency were tested to determine the effect which word frequency should have. The effect of word frequency was diminished by decreasing the word frequency scores until an optimum effect was observed. This is shown in Table L-1.

modification to frequency	rank 1	rank 2	rank 3	rank 4	rank 5	rank 6	rank 7	rank 8	rank 9	rank 10	rank 100
divide by 10	26.6%	35.8%	41.6%	45.6%	49.7%	53.1%	55.0%	56.7%	58.0%	60.0%	78.5%
divide by 100	26.6%	35.8%	41.6%	45.6%	49.7%	53.1%	55.0%	56.7%	58.0%	60.0%	78.5%
divide by 1000	26.3%	36.5%	42.4%	46.5%	50.1%	53.4%	55.5%	56.8%	58.2%	59.6%	78.0%
square root	27.8%	37.5%	43.8%	48.0%	52.5%	54.5%	56.5%	57.7%	59.0%	60.6%	78.5%
divide by 10 then take square root	27.8%	37.5%	43.8%	48.0%	52.5%	54.5%	56.5%	57.7%	59.0%	60.6%	78.5%
divide by 100 then take square root	27.8%	37.5%	43.8%	48.0%	52.5%	54.5%	56.5%	57.7%	59.0%	60.6%	78.5%
divide by 1000 then take square root	27.8%	38.5%	44.2%	47.9%	51.4%	53.3%	54.9%	56.0%	58.5%	59.8%	78.0%

Table L-1: The word level method combined with frequency information. Percent target word recognized at, or above, rank, with results given by modification to word frequency: 200 word data set.

Appendix M: Merging the Word Level Method with the Pattern Recognizer

The list of word alternatives generated by the pattern recognizer was merged with the list of words generated by the word level method (see section 6.4). A dummy confidence score was given to the words in the candidate list generated by the word level method. Table M-1 shows the effect of a range of values for this dummy confidence score.

confi-dence score	rank 1	rank 2	rank 3	rank 4	rank 5	rank 6	rank 7	rank 8	rank 9	rank 10	rank 100
100	63.4%	68.9%	71.5%	73.5%	74.9%	76.0%	77.0%	77.6%	78.6%	79.0%	90.0
99	63.4%	69.0%	71.6%	73.5%	74.9%	76.1%	77.0%	77.7%	78.6%	79.1%	90.0
98	63.5%	69.1%	71.6%	73.6%	75.0%	76.1%	77.1%	77.6%	78.6%	79.1%	89.9
97	63.6%	69.1%	71.7%	73.6%	75.0%	76.1%	77.1%	77.6%	78.6%	79.2%	89.9
96	63.6%	69.1%	71.7%	73.6%	75.0%	76.2%	77.1%	77.7%	78.7%	79.2%	89.9
95	63.6%	69.2%	71.7%	73.7%	75.1%	76.2%	77.1%	77.7%	78.7%	79.3%	89.9
94	63.7%	69.2%	71.8%	73.7%	75.1%	76.2%	77.2%	77.7%	78.6%	79.3%	89.9
93	63.7%	69.3%	71.9%	73.8%	75.1%	76.3%	77.2%	77.7%	78.6%	79.3%	89.9
92	63.8%	69.3%	71.9%	73.9%	75.1%	76.3%	77.2%	77.7%	78.7%	79.3%	89.9
91	63.8%	69.4%	71.9%	73.9%	75.2%	76.4%	77.2%	77.7%	78.7%	79.3%	89.9
90	63.9%	69.4%	72.0%	74.0%	75.3%	76.4%	77.2%	77.7%	78.7%	79.3%	89.9
89	63.9%	69.5%	72.2%	74.0%	75.3%	76.4%	77.2%	77.7%	78.7%	79.3%	89.9
88	63.9%	69.6%	72.2%	74.0%	75.2%	76.3%	77.2%	77.8%	78.8%	79.3%	89.9
87	63.9%	69.7%	72.2%	74.0%	75.2%	76.3%	77.2%	77.8%	78.8%	79.4%	89.9
86	64.0%	69.6%	72.2%	74.0%	75.2%	76.3%	77.2%	77.8%	78.8%	79.4%	89.9
85	64.0%	69.6%	72.2%	74.0%	75.3%	76.4%	77.2%	77.8%	78.8%	79.4%	89.9
84	64.1%	69.6%	72.2%	74.0%	75.3%	76.4%	77.2%	77.8%	78.8%	79.4%	89.9
83	64.1%	69.7%	72.2%	74.1%	75.4%	76.4%	77.2%	77.9%	78.8%	79.4%	89.9
82	64.1%	69.8%	72.3%	74.1%	75.3%	76.4%	77.1%	77.9%	78.8%	79.4%	89.9
81	64.1%	69.8%	72.3%	74.1%	75.3%	76.4%	77.1%	77.9%	78.9%	79.4%	89.9
80	64.1%	69.9%	72.2%	74.1%	75.3%	76.4%	77.1%	77.9%	78.9%	79.4%	89.9
79	64.1%	69.9%	72.3%	74.2%	75.3%	76.4%	77.1%	77.9%	78.9%	79.4%	89.9
78	64.1%	70.0%	72.3%	74.1%	75.3%	76.4%	77.2%	77.9%	78.9%	79.5%	89.9
77	64.1%	70.0%	72.3%	74.1%	75.3%	76.3%	77.2%	77.9%	79.0%	79.5%	89.9
76	64.2%	69.9%	72.3%	74.2%	75.3%	76.3%	77.2%	77.9%	78.9%	79.4%	89.9

confi-dence score	rank 1	rank 2	rank 3	rank 4	rank 5	rank 6	rank 7	rank 8	rank 9	rank 10	rank 100
75	64.2%	69.9%	72.4%	74.1%	75.3%	76.3%	77.2%	77.9%	78.9%	79.4%	89.9
74	64.2%	70.0%	72.4%	74.2%	75.4%	76.3%	77.2%	77.9%	78.9%	79.4%	89.9
73	64.3%	69.9%	72.4%	74.2%	75.4%	76.3%	77.2%	77.9%	78.9%	79.5%	89.9
72	64.3%	69.9%	72.4%	74.3%	75.4%	76.3%	77.1%	77.8%	78.9%	79.5%	89.9
71	64.3%	70.0%	72.5%	74.3%	75.4%	76.2%	77.2%	77.9%	78.9%	79.5%	89.9
70	64.4%	70.1%	72.5%	74.3%	75.4%	76.2%	77.2%	77.8%	78.9%	79.5%	89.9
69	64.4%	70.2%	72.5%	74.3%	75.4%	76.2%	77.1%	77.8%	78.9%	79.4%	89.9
68	64.4%	70.2%	72.5%	74.3%	75.4%	76.2%	77.2%	77.8%	78.7%	79.4%	89.9
67	64.4%	70.2%	72.5%	74.3%	75.4%	76.2%	77.2%	77.8%	78.7%	79.4%	89.9
66	64.5%	70.2%	72.5%	74.2%	75.3%	76.2%	77.2%	77.8%	78.8%	79.4%	89.9
65	64.5%	70.2%	72.5%	74.2%	75.4%	76.2%	77.2%	77.8%	78.8%	79.4%	89.9
64	64.5%	70.2%	72.6%	74.2%	75.3%	76.2%	77.2%	77.8%	78.8%	79.3%	89.9
63	64.6%	70.2%	72.6%	74.2%	75.3%	76.3%	77.2%	77.9%	78.7%	79.3%	89.9
62	64.6%	70.3%	72.6%	74.2%	75.3%	76.3%	77.3%	77.8%	78.7%	79.3%	89.9
61	64.6%	70.3%	72.5%	74.2%	75.3%	76.4%	77.3%	77.8%	78.7%	79.3%	89.9
60	64.6%	70.3%	72.6%	74.1%	75.3%	76.3%	77.2%	77.8%	78.7%	79.3%	89.9
59	64.6%	70.3%	72.6%	74.2%	75.4%	76.3%	77.2%	77.8%	78.7%	79.3%	89.9
58	64.6%	70.3%	72.6%	74.2%	75.3%	76.2%	77.2%	77.7%	78.6%	79.3%	89.9
57	64.6%	70.4%	72.6%	74.2%	75.3%	76.3%	77.2%	77.8%	78.6%	79.2%	89.9
56	64.6%	70.4%	72.6%	74.2%	75.4%	76.2%	77.2%	77.7%	78.6%	79.2%	89.9
55	64.6%	70.4%	72.6%	74.3%	75.3%	76.2%	77.2%	77.7%	78.6%	79.2%	89.9
54	64.7%	70.4%	72.6%	74.3%	75.3%	76.2%	77.2%	77.7%	78.6%	79.2%	89.9
53	64.7%	70.4%	72.7%	74.3%	75.3%	76.2%	77.2%	77.7%	78.6%	79.2%	89.9
52	64.8%	70.4%	72.7%	74.2%	75.3%	76.1%	77.2%	77.7%	78.6%	79.2%	89.9
51	64.9%	70.5%	72.7%	74.2%	75.2%	76.1%	77.1%	77.7%	78.6%	79.2%	89.9
50	64.9%	70.6%	72.7%	74.2%	75.2%	76.1%	77.0%	77.7%	78.6%	79.2%	89.9
49	65.0%	70.6%	72.8%	74.2%	75.1%	76.0%	77.0%	77.7%	78.6%	79.2%	89.9
48	65.0%	70.5%	72.8%	74.2%	75.1%	76.0%	77.0%	77.7%	78.6%	79.2%	89.9
47	65.0%	70.5%	72.7%	74.2%	75.2%	76.1%	77.0%	77.7%	78.6%	79.2%	89.9
46	65.0%	70.6%	72.7%	74.1%	75.1%	76.0%	77.0%	77.6%	78.6%	79.2%	89.9
45	65.0%	70.6%	72.8%	74.1%	75.1%	76.0%	76.9%	77.6%	78.5%	79.1%	89.9
44	65.0%	70.6%	72.8%	74.1%	75.1%	76.0%	77.0%	77.6%	78.6%	79.1%	89.9
43	65.1%	70.6%	72.8%	74.1%	75.1%	75.9%	77.0%	77.5%	78.5%	79.1%	89.9
42	65.1%	70.6%	72.8%	74.1%	75.0%	75.9%	76.9%	77.5%	78.4%	79.0%	89.9
41	65.2%	70.6%	72.8%	74.1%	75.0%	75.9%	76.8%	77.4%	78.4%	79.0%	89.9
40	65.2%	70.7%	72.8%	74.1%	75.0%	75.8%	76.8%	77.3%	78.4%	79.0%	89.9
39	65.3%	70.6%	72.8%	74.1%	75.0%	75.9%	76.7%	77.3%	78.4%	79.0%	89.9
38	65.2%	70.7%	72.8%	74.0%	75.0%	75.9%	76.7%	77.3%	78.3%	79.0%	89.9
37	65.3%	70.7%	72.8%	74.0%	75.0%	75.9%	76.7%	77.3%	78.3%	79.0%	89.9
36	65.2%	70.7%	72.8%	74.0%	75.0%	75.9%	76.7%	77.3%	78.3%	78.9%	89.9

confi-dence score	rank 1	rank 2	rank 3	rank 4	rank 5	rank 6	rank 7	rank 8	rank 9	rank 10	rank 100
35	65.2%	70.7%	72.8%	74.0%	75.0%	75.9%	76.7%	77.3%	78.3%	78.9%	89.9
34	65.1%	70.7%	72.7%	74.0%	75.0%	75.9%	76.7%	77.2%	78.3%	78.9%	89.9
33	65.2%	70.7%	72.8%	74.0%	75.0%	75.9%	76.7%	77.3%	78.3%	78.9%	89.9
32	65.1%	70.7%	72.8%	74.0%	75.0%	75.9%	76.7%	77.3%	78.2%	78.9%	89.9
31	65.1%	70.7%	72.7%	74.0%	74.9%	75.9%	76.7%	77.3%	78.2%	78.8%	89.9
30	65.2%	70.6%	72.8%	74.0%	75.0%	75.9%	76.7%	77.2%	78.2%	78.8%	89.9
29	65.1%	70.6%	72.7%	74.0%	75.0%	75.9%	76.6%	77.2%	78.1%	78.8%	89.9
28	65.1%	70.6%	72.8%	74.0%	74.9%	75.9%	76.6%	77.2%	78.2%	78.8%	89.9
27	65.1%	70.6%	72.7%	74.0%	74.9%	75.8%	76.6%	77.2%	78.1%	78.8%	89.9
26	65.0%	70.5%	72.7%	74.0%	74.9%	75.9%	76.6%	77.1%	78.1%	78.7%	89.9
25	65.1%	70.4%	72.6%	73.9%	74.9%	75.9%	76.6%	77.1%	78.1%	78.7%	89.9
24	65.1%	70.4%	72.6%	73.9%	74.9%	75.9%	76.6%	77.1%	78.1%	78.7%	89.9
23	65.0%	70.4%	72.6%	73.9%	74.9%	75.9%	76.6%	77.1%	78.1%	78.6%	89.9
22	65.0%	70.4%	72.6%	73.9%	75.0%	75.9%	76.6%	77.1%	78.1%	78.6%	89.9
21	64.9%	70.4%	72.6%	73.9%	75.0%	75.9%	76.6%	77.1%	78.1%	78.6%	89.9
20	64.9%	70.5%	72.6%	73.9%	75.0%	75.9%	76.6%	77.1%	78.1%	78.6%	89.9
19	64.9%	70.5%	72.5%	73.9%	75.0%	75.9%	76.6%	77.1%	78.1%	78.7%	89.9
18	64.8%	70.5%	72.5%	73.9%	74.9%	75.8%	76.6%	77.1%	78.1%	78.7%	89.9
17	64.8%	70.5%	72.5%	73.9%	74.9%	75.8%	76.6%	77.1%	78.1%	78.7%	89.9
16	64.8%	70.5%	72.5%	73.9%	74.9%	75.8%	76.6%	77.1%	78.1%	78.7%	89.9
15	64.8%	70.4%	72.5%	73.9%	74.9%	75.8%	76.6%	77.2%	78.0%	78.6%	89.9
14	64.8%	70.4%	72.5%	73.9%	74.9%	75.8%	76.6%	77.1%	78.0%	78.6%	89.9
13	64.9%	70.4%	72.5%	73.9%	74.9%	75.8%	76.6%	77.2%	78.0%	78.6%	89.9
12	64.9%	70.4%	72.4%	73.9%	74.8%	75.7%	76.6%	77.2%	78.0%	78.6%	89.9
11	64.8%	70.4%	72.4%	73.9%	74.8%	75.7%	76.6%	77.2%	78.0%	78.6%	89.9
10	64.9%	70.4%	72.4%	73.9%	74.9%	75.7%	76.6%	77.2%	78.0%	78.6%	89.9
9	64.9%	70.3%	72.4%	73.9%	74.9%	75.7%	76.6%	77.2%	78.0%	78.6%	89.9
8	64.8%	70.2%	72.4%	73.9%	74.9%	75.7%	76.6%	77.2%	78.0%	78.6%	89.9
7	64.8%	70.2%	72.3%	73.9%	74.9%	75.7%	76.6%	77.2%	78.0%	78.6%	89.9
6	64.7%	70.2%	72.2%	73.9%	74.9%	75.7%	76.6%	77.2%	78.0%	78.6%	89.9
5	64.7%	70.2%	72.2%	73.9%	74.9%	75.7%	76.6%	77.2%	78.0%	78.6%	89.9
4	64.6%	70.1%	72.1%	73.7%	74.8%	75.7%	76.6%	77.2%	78.0%	78.6%	89.9
3	64.6%	70.0%	72.1%	73.7%	74.8%	75.7%	76.6%	77.2%	78.0%	78.6%	89.9
2	64.5%	70.0%	72.1%	73.6%	74.8%	75.7%	76.6%	77.1%	78.0%	78.6%	89.9
1	64.4%	69.9%	72.2%	73.6%	74.8%	75.7%	76.6%	77.1%	78.0%	78.6%	89.9
0	64.3%	69.7%	72.2%	73.7%	74.8%	75.6%	76.6%	77.1%	78.0%	78.6%	89.9

Table M-1: Merging the word level method with the pattern recognizer. Percent target word recognized at, or above, rank, results given by confidence score: 200 word data set.

Appendix N: Letter Confusion Matrices for the Word Level Method

Table N-1 and Table N-2 show the letter confusion matrices for the word level method (see section 6.2). Size considerations mean that each matrix has been split into two. The confusion matrices were constructed by comparing the cues derived from the training set with their target values. The percentage value is a measure of confidence: common confusions have higher scores. The matrices are shown from confusion to target. This is the way that they are used in practice. The matrices have been left blank where no confusion exists. A dash has been placed in the matrices whenever a letter coincides with itself. The confusion matrices are not symmetrical. This is also the case for human readers [Bouma, 1971]. Note that the confusion matrices given here differ from those shown in Appendix H. The reason for this is that the two sets of matrices were constructed from different data. The matrices in Appendix H were constructed using the candidates from the complete data set.

In the approach described in Chapter 4 the percentage values are not used for any calculations (see especially section 4.4.2)). Instead, the percentage values are used to rank the letter confusions from the most likely to the least likely confusion.

	a	b	c	d	e	f	g	h	i	j	k	l	m
a	-		14%	14%	17%		1%		7%		1%	2%	5%
b		-									33%		
c	21%	3%	-	8%	18%	3%	3%	3%	3%				
d				-	67%								
e					-			20%					
f	3%	1%	3%	2%	7%	-	3%	4%	3%	9%	1%	1%	1%
g							-						
h	5%					5%		-				5%	5%
i	5%				5%	5%		14%	-				9%
j										-			
k											-		
l			3%		14%	11%	3%	3%	9%	3%		-	
m									40%				-
n	3%				1%			5%	5%				19%
o						20%							40%
p		11%				11%							11%
q							100%						
r	7%			4%	14%	7%	4%		4%			7%	
s		14%				21%			7%				
t	8%			3%		30%		5%	14%	5%			3%
u	33%				33%								33%
v													
w	12%				6%			6%	9%	3%			15%
x													
y													
z													

	n	o	p	q	r	s	t	u	v	w	x	y	z
a	2%	14%		4%	1%	1%	1%	7%	2%	6%			1%
b						33%							33%
c		5%		8%	3%	13%	3%	3%	3%	5%			
d									33%				
e	40%		20%			20%							
f	7%	3%	2%	1%	4%	5%	25%	3%	1%	3%		9%	1%
g				100%									
h	5%				10%		40%	15%		10%			
i	5%	18%	5%		9%	5%	9%	9%				5%	
j												100%	
k													
l	6%	9%				9%	14%	3%		9%		6%	
m			20%				20%			20%			
n	-	1%	2%	2%	7%		1%	29%	11%	16%			
o	20%	-			20%								
p	33%		-		22%								11%
q				-									
r	11%		4%		-	14%		4%	18%			4%	
s	14%		7%		14%	-							21%
t	3%		5%			5%	-	3%		8%		3%	5%
u								-					
v								100%	-				
w	6%	3%		3%	3%		6%	15%	12%	-			
x											-		
y												-	
z													-

Table N-1: Word level method: confusion matrix for first letter

	a	b	c	d	e	f	g	h	i	j	k	l	m
a	-												
b		-											
c			-	100%									
d	9%			-	11%							20%	
e	3%		4%	13%	-	9%		1%				12%	
f					50%	-							
g							-						
h					100 %			-					
i									-				
j										-			
k											-		
l				8%	3%	5%	3%					-	
m													-
n	1%		1%	1%	3%	1%		5%				6%	25%
o					67%								
p													
q													
r			2%	6%	15%	2%		2%				6%	7%
s	4%		2%	2%	8%		8%	2%				2%	8%
t			4%	4%	7%	7%	4%					11%	
u													
v													
w													
x													
y					3%		31%					6%	3%
z													

	n	o	p	q	r	s	t	u	v	w	x	y	z
a													
b													
c													
d	9%				6%		37%				3%	6%	
e	1%	12%			18%	3%	16%	2%		1%	3%	1%	
f	50%												
g						18%	9%					73%	
h													
i													
j													
k													
l	39%	5%			3%	8%	18%					8%	
m													
n	-	5%			18%	10%	8%	5%		8%	1%	2%	
o	33%	-											
p			-										
q				-									
r	39%	2%			-	4%	11%			4%		2%	
s	10%	6%			6%	-	10%				4%	28%	
t	15%				7%	19%	-				11%	11%	
u								-					
v									-				
w										-			
x											-		
y	6%				6%	39%	8%					-	
z													-

Table N-2: Word level method: confusion matrix for last letter

Appendix O: Letter Verification Applied to the Word Level Method

The letter verification recognizer was applied to the word level method merged with the pattern recognizer (see section 6.5). The score obtained from letter verification was multiplied by a constant. Table O-1 shows the effect of a range of values for this constant.

constant	rank 1	rank 2	rank 3	rank 4	rank 5	rank 6	rank 7	rank 8	rank 9	rank 10	rank 100
1.0	69.4%	75.5%	77.8%	79.4%	80.4%	81.1%	81.7%	82.1%	82.8%	83.4%	89.9%
0.99	69.5%	75.5%	77.8%	79.4%	80.4%	81.1%	81.7%	82.1%	82.9%	83.4%	89.9%
0.98	69.5%	75.5%	77.8%	79.4%	80.4%	81.0%	81.7%	82.1%	82.9%	83.4%	89.9%
0.97	69.5%	75.5%	77.7%	79.4%	80.4%	81.0%	81.7%	82.1%	82.9%	83.4%	89.9%
0.96	69.5%	75.5%	77.8%	79.3%	80.4%	81.0%	81.6%	82.1%	82.8%	83.4%	89.9%
0.95	69.4%	75.5%	77.7%	79.3%	80.4%	81.0%	81.6%	82.1%	82.8%	83.4%	89.9%
0.9	69.4%	75.5%	77.8%	79.3%	80.5%	81.0%	81.6%	82.1%	82.7%	83.4%	89.9%
0.9	69.4%	75.5%	77.7%	79.3%	80.4%	81.0%	81.6%	82.1%	82.7%	83.4%	89.9%
0.9	69.5%	75.5%	77.7%	79.2%	80.4%	81.0%	81.6%	82.1%	82.7%	83.4%	89.9%
0.9	69.4%	75.4%	77.7%	79.3%	80.4%	81.0%	81.6%	82.1%	82.8%	83.4%	89.9%
0.9	69.4%	75.5%	77.7%	79.3%	80.4%	81.1%	81.6%	82.1%	82.7%	83.4%	89.9%
0.9	69.5%	75.4%	77.8%	79.3%	80.4%	81.0%	81.6%	82.0%	82.8%	83.4%	89.9%
0.9	69.5%	75.5%	77.7%	79.3%	80.4%	81.1%	81.6%	82.0%	82.8%	83.4%	89.9%
0.9	69.5%	75.5%	77.8%	79.3%	80.4%	81.1%	81.6%	82.0%	82.8%	83.4%	89.9%
0.9	69.4%	75.5%	77.7%	79.3%	80.4%	81.0%	81.6%	82.0%	82.8%	83.4%	89.9%
0.9	69.4%	75.5%	77.7%	79.2%	80.4%	81.0%	81.5%	82.0%	82.8%	83.4%	89.9%
0.8	69.5%	75.5%	77.7%	79.3%	80.3%	81.0%	81.6%	82.0%	82.8%	83.4%	89.9%
0.8	69.5%	75.5%	77.7%	79.2%	80.4%	81.0%	81.5%	82.1%	82.8%	83.3%	89.9%
0.8	69.5%	75.5%	77.7%	79.2%	80.3%	81.0%	81.5%	82.0%	82.8%	83.3%	89.9%
0.8	69.6%	75.5%	77.6%	79.2%	80.3%	81.0%	81.5%	82.0%	82.8%	83.3%	89.9%
0.8	69.4%	75.6%	77.6%	79.2%	80.3%	81.0%	81.5%	82.0%	82.8%	83.3%	89.9%
0.8	69.6%	75.6%	77.6%	79.2%	80.3%	80.9%	81.6%	82.0%	82.8%	83.3%	89.9%
0.8	69.6%	75.6%	77.6%	79.1%	80.2%	80.9%	81.5%	81.9%	82.7%	83.3%	89.9%
0.8	69.6%	75.6%	77.6%	79.1%	80.3%	80.9%	81.5%	81.9%	82.7%	83.2%	89.9%

constant	rank 1	rank 2	rank 3	rank 4	rank 5	rank 6	rank 7	rank 8	rank 9	rank 10	rank 100
0.8	69.6%	75.6%	77.6%	79.1%	80.2%	80.8%	81.5%	81.9%	82.7%	83.2%	89.9%
0.8	69.5%	75.6%	77.6%	79.1%	80.2%	80.8%	81.4%	81.9%	82.7%	83.2%	89.9%
0.7	69.5%	75.6%	77.6%	79.2%	80.2%	80.7%	81.4%	81.8%	82.6%	83.2%	89.9%
0.7	69.5%	75.6%	77.6%	79.1%	80.1%	80.7%	81.4%	81.8%	82.6%	83.2%	89.9%
0.7	69.5%	75.6%	77.5%	79.0%	80.0%	80.7%	81.4%	81.9%	82.6%	83.1%	89.9%
0.7	69.5%	75.5%	77.6%	79.0%	80.1%	80.7%	81.4%	81.8%	82.5%	83.2%	89.9%
0.7	69.5%	75.6%	77.6%	79.0%	80.0%	80.7%	81.4%	81.8%	82.5%	83.1%	89.9%
0.7	69.4%	75.6%	77.6%	79.0%	80.0%	80.7%	81.4%	81.8%	82.5%	83.1%	89.9%
0.7	69.4%	75.5%	77.6%	79.0%	80.0%	80.7%	81.3%	81.8%	82.5%	83.1%	89.9%
0.7	69.3%	75.5%	77.6%	79.0%	79.9%	80.7%	81.3%	81.8%	82.5%	83.1%	89.9%
0.7	69.3%	75.5%	77.6%	79.0%	79.9%	80.7%	81.2%	81.7%	82.5%	83.1%	89.9%
0.7	69.3%	75.5%	77.6%	78.9%	79.9%	80.7%	81.2%	81.8%	82.4%	83.1%	89.9%
0.6	69.3%	75.5%	77.6%	79.0%	79.9%	80.7%	81.2%	81.8%	82.5%	83.1%	89.9%
0.6	69.2%	75.5%	77.6%	79.0%	79.9%	80.7%	81.2%	81.8%	82.4%	83.0%	89.9%
0.6	69.2%	75.5%	77.5%	79.0%	79.9%	80.6%	81.2%	81.8%	82.4%	83.0%	89.9%
0.6	69.1%	75.6%	77.5%	79.0%	79.9%	80.6%	81.1%	81.8%	82.4%	82.9%	89.9%
0.6	69.2%	75.5%	77.5%	78.9%	79.8%	80.6%	81.1%	81.8%	82.4%	82.9%	89.9%
0.6	69.1%	75.4%	77.5%	79.0%	79.8%	80.5%	81.1%	81.8%	82.4%	82.9%	89.9%
0.6	69.1%	75.4%	77.4%	79.0%	79.8%	80.5%	81.1%	81.8%	82.4%	83.0%	89.9%
0.6	69.1%	75.5%	77.4%	79.0%	79.7%	80.5%	81.1%	81.8%	82.4%	82.9%	89.9%
0.6	69.1%	75.4%	77.4%	78.9%	79.7%	80.4%	81.1%	81.8%	82.4%	82.9%	89.9%
0.6	69.2%	75.5%	77.4%	78.8%	79.7%	80.5%	81.1%	81.7%	82.4%	82.9%	89.9%
0.5	69.3%	75.4%	77.4%	78.8%	79.6%	80.4%	80.9%	81.7%	82.4%	82.9%	89.9%
0.5	69.2%	75.3%	77.4%	78.8%	79.7%	80.4%	80.9%	81.6%	82.4%	82.9%	89.9%
0.5	69.2%	75.3%	77.4%	78.8%	79.7%	80.4%	80.9%	81.6%	82.4%	82.9%	89.9%
0.5	69.3%	75.2%	77.4%	78.8%	79.7%	80.4%	80.9%	81.6%	82.3%	82.9%	89.9%
0.5	69.2%	75.2%	77.4%	78.7%	79.7%	80.4%	80.9%	81.6%	82.3%	82.9%	89.9%
0.5	69.2%	75.2%	77.4%	78.7%	79.7%	80.3%	80.9%	81.6%	82.3%	82.8%	89.9%
0.5	69.2%	75.2%	77.4%	78.7%	79.6%	80.3%	80.9%	81.6%	82.3%	82.8%	89.9%
0.5	69.3%	75.2%	77.3%	78.7%	79.6%	80.3%	80.9%	81.6%	82.2%	82.8%	89.9%
0.5	69.2%	75.1%	77.3%	78.5%	79.6%	80.3%	81.0%	81.6%	82.2%	82.7%	89.9%
0.5	69.3%	75.2%	77.3%	78.5%	79.6%	80.2%	80.9%	81.6%	82.2%	82.7%	89.9%
0.4	69.2%	75.1%	77.3%	78.5%	79.6%	80.2%	80.9%	81.6%	82.2%	82.7%	89.9%
0.4	69.1%	75.1%	77.2%	78.6%	79.6%	80.2%	81.0%	81.6%	82.2%	82.7%	89.9%
0.4	69.1%	75.1%	77.2%	78.5%	79.6%	80.3%	80.9%	81.6%	82.2%	82.7%	89.9%
0.4	69.1%	75.1%	77.2%	78.5%	79.6%	80.3%	80.9%	81.6%	82.2%	82.7%	89.9%
0.4	68.9%	75.1%	77.2%	78.4%	79.6%	80.2%	80.8%	81.6%	82.2%	82.7%	89.9%
0.4	68.9%	75.0%	77.3%	78.4%	79.6%	80.3%	80.9%	81.6%	82.1%	82.6%	89.9%
0.4	68.9%	74.9%	77.2%	78.3%	79.5%	80.1%	80.9%	81.5%	82.2%	82.6%	89.9%
0.4	68.7%	75.0%	77.2%	78.3%	79.5%	80.1%	80.8%	81.4%	82.2%	82.6%	89.9%
0.4	68.7%	75.0%	77.1%	78.3%	79.5%	80.1%	80.8%	81.4%	82.1%	82.6%	89.9%
0.4	68.7%	74.9%	77.0%	78.3%	79.4%	80.0%	80.8%	81.4%	82.1%	82.6%	89.9%

constant	rank 1	rank 2	rank 3	rank 4	rank 5	rank 6	rank 7	rank 8	rank 9	rank 10	rank 100
0.3	68.6%	74.9%	76.8%	78.3%	79.4%	80.0%	80.8%	81.4%	82.2%	82.6%	89.9%
0.3	68.6%	74.8%	76.7%	78.3%	79.4%	80.0%	80.8%	81.4%	82.1%	82.6%	89.9%
0.3	68.4%	74.6%	76.6%	78.3%	79.4%	79.9%	80.7%	81.4%	82.1%	82.5%	89.9%
0.3	68.4%	74.5%	76.6%	78.1%	79.3%	79.8%	80.7%	81.3%	82.1%	82.6%	89.9%
0.3	68.4%	74.3%	76.5%	78.1%	79.1%	79.9%	80.5%	81.2%	82.1%	82.5%	89.9%
0.3	68.2%	74.1%	76.4%	78.0%	79.0%	79.8%	80.5%	81.2%	82.1%	82.5%	89.9%
0.3	68.2%	74.1%	76.3%	77.9%	78.8%	79.8%	80.5%	81.1%	82.1%	82.5%	89.9%
0.3	68.1%	73.9%	76.1%	77.7%	78.8%	79.8%	80.5%	81.1%	82.0%	82.5%	89.9%
0.3	68.1%	73.8%	76.0%	77.6%	78.6%	79.6%	80.5%	81.1%	82.0%	82.5%	89.9%
0.3	68.0%	73.6%	75.8%	77.2%	78.4%	79.4%	80.5%	81.0%	81.9%	82.3%	89.9%
0.2	67.9%	73.6%	75.7%	77.1%	78.3%	79.3%	80.4%	81.0%	81.9%	82.3%	89.9%
0.2	67.7%	73.5%	75.6%	77.1%	78.2%	79.2%	80.3%	81.0%	81.9%	82.3%	89.9%
0.2	67.6%	73.3%	75.5%	76.9%	77.9%	78.9%	80.2%	80.9%	81.8%	82.2%	89.9%
0.2	67.5%	73.2%	75.4%	76.7%	77.7%	78.9%	80.0%	80.9%	81.7%	82.3%	89.9%
0.2	67.5%	73.1%	75.2%	76.6%	77.6%	78.6%	79.9%	80.9%	81.7%	82.3%	89.9%
0.2	67.3%	72.9%	75.0%	76.5%	77.4%	78.5%	79.6%	80.8%	81.7%	82.2%	89.9%
0.2	67.4%	72.9%	74.9%	76.2%	77.2%	78.4%	79.6%	80.6%	81.6%	82.1%	89.9%
0.2	67.2%	72.8%	74.8%	76.0%	77.0%	78.1%	79.2%	80.4%	81.4%	82.0%	89.9%
0.2	67.2%	72.6%	74.8%	75.9%	77.0%	77.7%	79.1%	80.3%	81.3%	82.0%	89.9%
0.2	67.0%	72.5%	74.7%	75.8%	76.7%	77.6%	78.8%	80.0%	81.1%	81.9%	89.9%
0.1	66.9%	72.4%	74.6%	75.7%	76.6%	77.4%	78.5%	79.7%	81.0%	81.7%	89.9%
0.1	66.7%	72.3%	74.4%	75.6%	76.6%	77.3%	78.1%	79.4%	80.7%	81.6%	89.9%
0.1	66.7%	72.2%	74.3%	75.5%	76.5%	77.1%	78.0%	79.1%	80.5%	81.6%	89.9%
0.1	66.5%	72.0%	74.1%	75.5%	76.3%	77.1%	77.8%	78.9%	80.1%	81.2%	89.9%
0.1	66.2%	71.8%	73.9%	75.3%	76.2%	77.0%	77.7%	78.6%	79.8%	80.9%	89.9%
0.9	66.3%	71.7%	73.8%	75.0%	76.1%	77.0%	77.7%	78.5%	79.5%	80.5%	89.9%
0.8	66.2%	71.4%	73.6%	74.8%	76.0%	76.7%	77.5%	78.4%	79.2%	80.1%	89.9%
0.7	66.0%	71.3%	73.3%	74.7%	75.7%	76.6%	77.5%	78.2%	79.0%	79.9%	89.9%
0.6	65.9%	71.2%	73.3%	74.6%	75.6%	76.5%	77.3%	78.0%	78.8%	79.7%	89.9%
0.5	65.7%	71.0%	73.2%	74.3%	75.6%	76.3%	77.0%	77.9%	78.7%	79.3%	89.9%
0.4	65.6%	71.0%	73.1%	74.3%	75.3%	76.2%	76.9%	77.7%	78.6%	79.2%	89.9%
0.3	65.3%	70.8%	72.9%	74.1%	75.1%	76.0%	76.8%	77.5%	78.5%	79.1%	89.9%
0.2	65.3%	70.7%	72.9%	74.1%	75.0%	75.9%	76.8%	77.3%	78.4%	79.0%	89.9%
0.1	65.3%	70.7%	72.8%	74.0%	75.0%	75.9%	76.7%	77.3%	78.3%	79.0%	89.9%

Table O-1: Merging the word level method with the letter verification recognizer. Percent target word recognized at, or above, rank, results given by constant applied to the letter verification recognizer: 200 word data set.

The Integration of Knowledge Sources for Word Recognition

G.J. Bellaby & L.J. Evett

L.J. Evett and T.G. Rose (Eds.), Computational linguistics for speech and handwriting recognition, One day workshop in the AISB 1994 Workshop series, Leeds, April, 1994.

ABSTRACT

Script recognition systems require the use of context to disambiguate input fully. There is a limit on performance beyond which further development at the pattern level will not improve performance significantly. The paper looks at the integration of pattern level, word level and meta-word level information to produce an efficient and robust recognition system. The main topic is the application of information at the word level. Work on the integration of information taken from the meta-word level (semantic and syntactic) is underway. Experiment 1 shows that the ability of a pattern recognition system to unambiguously identify cursive handwriting can actually decrease as the software is trained on new samples of handwriting. Experiment 2 compares the performance of human readers and a machine system. Human readers show a much greater effect of word context than the machine. It is proposed that the factors producing this superior context effect for human readers be built into the machine system. The arguments presented are also relevant to speech recognition.

1 INTRODUCTION

There is great interest in the development of a machine interface which can use a natural means of communication. One standard mode of communication is writing. However, cursive script is the usual form of handwriting. Cursive script recognition is problematic because of the great variability between writers and in the writing of a single individual as well as difficulties in segmenting characters.

It is apparent from examining the way in which humans read that context effects are very strong and that one particular and important point where contextual information is used is at the word level. One way of depicting the reading process is to make a distinction between the letter level, the word level and the meta-word level and to consider the different sources of information which can be utilized at each of these levels. This can be seen in Figure 1. This paper presents some results arising from work concerned with the application of information at the word level. A number of different sources of information, such as word frequency and word length, have been chosen as being of use at the word level. This selection was made on the basis of research carried out within the field of cognitive psychology [e.g. Lindsay & Norman, 1977; DeZuniga et.al, 1991; Morton, 1969].

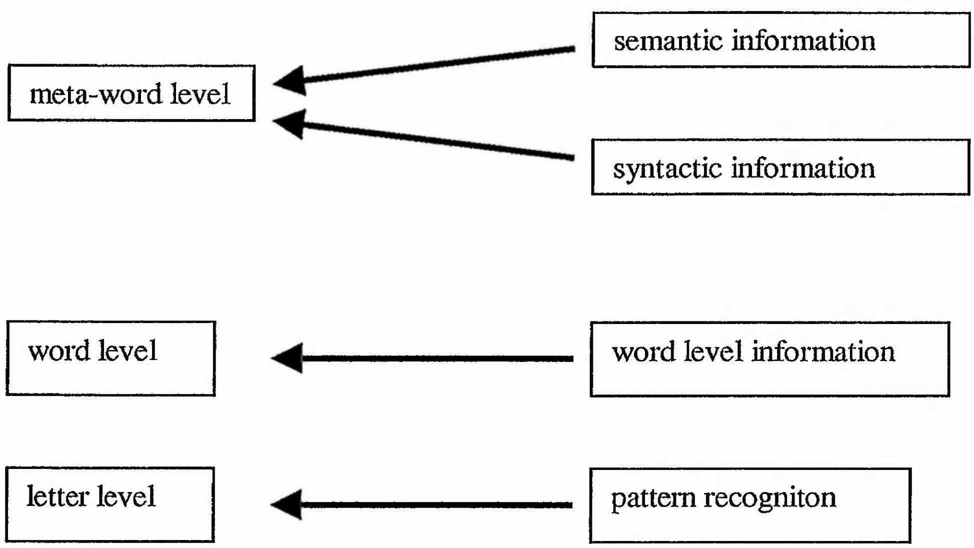


Figure 1: One view of the reading process

The reason for examining the way in which humans read is that people are the only efficient readers of unconstrained handwritten text. It is reasonable to use information about human reading processes and the sources of information which they use during these processes because reading is a human activity. Handwriting has developed with a human audience in mind. Given that a machine system is expected to read normal handwriting, it seems a well-grounded assumption that such a system will have to rely on the same sources of information as its human counterparts.

The particular topic of this paper is the application of information at the word level. However, a second area of interest is the integration of the different aspects of the reading process. It is hoped that by integrating the different aspects of the reading process that a stronger, more robust, recognition system can be developed. In particular, one which is capable of dealing with the particular problems of unconstrained cursive handwriting and a large lexicon. The word level is a good area to begin the process of integration standing as it does between the letter level and the meta-word level. The word level is an intermediary between the pattern recognition side of script recognition and the language side. The word level can be seen to partake of both of these two different kinds of information. The word level utilizes information which is derived from pattern recognition (e.g. the exploitation of features) and information which has a broader contextual foundation (e.g. lexical information and word frequency information).

The word level can be considered the first point within the reading process that contextual information is applied. A movement from the contextual considerations of the word level to the broader contextual constraints of syntax and semantics is both a reasonable move and methodologically sound. Those areas where the different levels of the reading process overlap are important for investigating the ways in which the various sources of information can be integrated in an effective and efficient manner.

2 EXPERIMENT 1: IMPROVING MACHINE PERFORMANCE

A central argument of this paper is that natural cursive script is inherently ambiguous and that a machine system cannot be expected to recognize individual letters beyond a limited degree of accuracy. It is the case that the same written pattern can represent more than one word. For example, the word given in Figure 2 can be read as "dog", but it could also signify the word "clog" or even "cloy". In other words, this pattern of lines on the page can legitimately signify three different words.

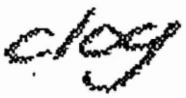


Figure 2: A handwriting example

It is only when contextual factors are added to the pattern recognition process that the machine recognition of unconstrained handwriting will approximate that of human recognition. Very little data on how well the pattern recognition stage of a script recognition system should perform has been accumulated. However, human readers do not recognize individual letters perfectly [Suen, 1983]. The reason for this is that human handwriting displays great variability. It is also ambiguous. Characters are just a set of arbitrary strokes whose variability between people, and even within the handwriting of one person, can be substantial. The English character set and writing conventions compensate for this, at least in part, by means of the duplication of information, or the use of supporting data. However, it is not actually necessary for people to recognize every individual letter or word in order to read a text successfully. Human readers use word, syntactic and semantic information, as well as letter recognition, in their recognition of handwriting. The existence of ambiguity in handwriting means that an opportunity exists to exploit contextual information.

It is often assumed by researchers working within the field of the machine recognition of handwriting that the recognition of individual letters can and should be improved indefinitely. A common approach is to develop a system whose recognition of letters is extremely accurate [Tappert, 1982]. Typically such systems are writer dependent, i.e. the recognition system is trained to recognize the writing of a particular user, or small group of users.

One way of improving recognition is to train the system on samples of handwriting. However, valid training will increase ambiguity. Therefore, the consequence of training will be that the ability of a pattern recognition system to unambiguously identify cursive script will decrease. This experiment is designed to demonstrate this effect. Such an effect would show that it is futile to attempt to improve performance indefinately.

The data used in this, and other, experiments was derived from the handwriting recognition system developed within the Nottingham Trent University. Ambiguity in handwritten words is real and valid. The handwriting recognition system works by matching letter and segmentation patterns to a pattern database. Hence it is in the database that ambiguity is represented.

The handwriting recognition system has a pattern recognizer that produces a set of individual characters, together with segmentation information about these characters, and basic information about the segmentation of words. These characters are combined to produce letter strings which are then filtered according to a lexicon to produce a set of word candidates. The resulting list of words is ranked according to a confidence score which has been given by the recognition system to each of the word alternatives. The maximum possible number of words in the list is limited to 100.

The experiment consisted of lower case cursive handwriting being presented to the pattern recognition software. A 15k lexicon was used. A sample of writing from 18 subjects was used, each of whom wrote down the same 200 words. The samples were of normal, clear cursive handwriting.The pattern recognition software had already been trained on other examples of handwriting from one third of these 18 subjects. The performance of the pattern recognition software on all of the samples was then recorded.

Three criteria were used to measure the performance of the recognition software: 1) best rate: the pattern recognition software placed the target word at the top of its list of alternatives. 2) rest rate: the recognition software gave the target word as an alternative but did not rank it top. 3) error rate: the recognition software completely failed to give the target word as an alternative.

The subjects were ranked according to the error rate produced by the recognition software on their samples of writing. On the basis of the average error rate, the 5 middle subjects were selected for training. This was because the system would have required limited training on writers with a low error rate and because it excluded poor handwriting. The effect of increased ambiguity upon a range of handwriting could therefore be observed. The pattern recognition software was then trained on the 200 word data samples for each of the 5 selected writers.

It was necessary to ensure that any ambiguity introduced by the training process was the result of real ambiguity in the subject's writing and not a consequence of simply increasing the size of the letter-segmentation database. It was therefore important that the size of the database was not arbitrarily increased, e.g. by simply adding new letter-segmentation patterns, but rather that any new patterns introduced were a direct consequence of the variability of actual letter forms.

The software was trained to recognize all of those words in the data samples which it was felt that a human reader would reasonably recognize. The balance of judgement in those cases of poor handwriting was always to leave the software untrained on the particular word.

Only those letter-segmentation patterns which corresponded to a real letter were chosen. More than one valid letter-segmentation pattern can exist for a particular letter in a given word. In this case only one letter-segmentation pattern per letter was chosen for the purpose of training. This helped forestall the problem of introducing false letter-segmentation patterns into the database. In this manner only the smallest possible number of letter-segmentation patterns were added to the database.

The methodology adopted can be shown by means of an example. Suppose that during the training process the recognition software was presented with the word "cat" and none of the letters in the word was at that point recognized by the pattern recognition software. If, during the course of training, only three letter-segmentation patterns were added to the database (one for "c", one for "a" and one for "t"), and if the addition of these three letter-segmentation patterns had the consequence that the word was now recognized by the software, then it was assumed that only legitimate letter- segmentations had been added to the database.

The performance of the pattern recognition software on all of the samples was again recorded. A comparison between the original and the new performance is given in Figure 3. Three significant effects of the training were noticeable. Firstly, the error rate (as should be expected) for the 5 samples of writing used in the training process significantly decreased, and the best rate significantly increased. (Sign test; $p < 0.031$.) Secondly, the error rate for all of the other 13 subjects also decreased and the best rate increased. Both of these changes were highly significant. (Sign test; $p < 0.001$.)

The third effect was not immediately apparent from the performance of the pattern recognition software but became clear as the data was further analysed. For every one of the 18 writers, including those 5 writers used in the training process, a number of words which were originally top ranked by the software were lowered in rank. In other words, the process of training software caused a highly significant movement from best to rest. (Sign test; $p < 0.001$.) The extent of this movement is shown in Figure 4.

What these results show is that the ability of a pattern recognition system to unambiguously identify cursive handwriting can indeed decrease as the software is trained. It might be suggested that these results are a result of the characteristics of the particular pattern recognition software used. However, it is difficult to see how any method of encoding and deciphering will fail to exhibit behaviour similar to that shown here. It is common for some letters to be written in a very similar fashion so that, for instance, the next to last letter of the word shown in Figure 5 may be an "r" but it could also be a "v". Ambiguity will increase as a machine recognition system is trained on such examples. Human readers will be confused when presented with an 'r' which resembles a 'v'. This confusion is a direct result of their need and ability to understand many different kinds of handwriting. It should therefore be accepted that a machine recognition system will experience similar difficulties to those experienced by human readers.

Secondly, these results demonstrate a way in which researchers within the field of the machine recognition of handwriting can be deceived with regard to the effectiveness of pattern recognition. It is all too easy to observe the best rate, and overall recognition rate, of a system increase whilst missing this small, but steady, increase in ambiguity. This effect will be less appreciable on writer dependent systems, and it may be masked by the use of small lexicons or limited training periods, but it will be present.

3 EXPERIMENT 2: THE WORD SUPERIORITY EFFECT

One great advantage which human readers have over a machine system which relies purely on pattern recognition is their ability to take into account contextual information at the word level. For example, it may be the case that words facilitate the recognition of their constituent letters by activation being fed back from the word-percept level to the letter level thus improving the perception of individual letters [McClelland et al, 1992]. A system with the ability to use context will be more efficient and flexible than one which simply uses pattern recognition.

Nine sources of information have been selected as being applicable to the word level: lexical, word frequency, word length, first letter, last letter, and the presence or absence of ascenders, descenders, i- dots and j-dots, and lastly t-crosses and f-crosses. No additions to the pattern recognition software have yet been made in order to exploit these particular sources of information. However, information about these features can be derived from the list of word alternatives given by the existing software. For example, a reasonable indication of the length of the target word can be found by calculating the mean length of the words appearing in the word list.

Leaving aside word frequency information for the moment, it is possible to show how useful these kinds of information can be within the recognition process and to indicate how they can be combined to good effect at the word level. Firstly, if it were possible to gain completely accurate information about the various word level features then it becomes significantly easier to select the target word as the most likely candidate from a list of alternatives. In those cases where the recognition software gives the target word as an alternative it is almost always possible to place the target word at the top of the list of alternatives.

Far more significantly, the use of word level features makes it viable to select new candidate words from the lexicon to add to the list of alternatives. This has particular relevance to those situations in which the recognition software completely fails to give the target word as an alternative. This information can be surprisingly efficient. For example, a 15k lexicon based on word frequency and containing every morphological variant of each of the words is used. Given exact information about the length of the word, its first letter, its last letter and whether ascenders, descenders, i-dots, j- dots, t- crosses and f-crosses are present or absent in the target word, then at the very worst a list only 12 words long would be selected from the lexicon. Even with the less accurate figures which have been derived from an examination of the list of alternatives suggested it is rare to produce a list which is unmanageably long.

The importance of such additions is that it now becomes possible to use syntactic and semantic information to make a selection from the alternatives. It can be suggested that the method of using word features to derive a list of possibilities is analogous to the human reader's ability to derive suitable candidates from even poor handwriting. More work is currently underway on expanding the contribution of these word level features to the recognition process.

One instance of the importance of context at the word level is the word superiority effect [Cattell, 1886; Baron & Thurston, 1973]. This is the effect noticeable with human readers whereby letters in words will probably be recognized more easily than letters in isolation or letters in non-words. This effect is well documented although no explanation for the effect has been agreed upon. However, word frequency can be shown to make a major contribution to the word superiority effect. Incidentally, the reason why most readers will tend to interpret the word in Figure 2 as "dog" rather than "clog" or "cloy" is that the first of these words has a much higher word frequency than the others.

An experiment was undertaken in order to suggest some figures for how well human readers recognize whole words and letters taken from these words. A comparison was made with the existing pattern recognition software, that is to say, without the addition of any word level information, apart from lexical, to the recognition process. This experiment demonstrates a standard effect in the human word recognition literature, but using cursive script rather than printed characters.

The experiment consisted of reasonably clear lower case cursive handwriting being presented to human subjects. Two connected sets of data were used: whole words and letters taken from these words. The subjects were asked to recognize what letters or words they had seen. Only one response to an image by the subjects was allowed. The results were compared with the performance of the handwriting recognition software.

Figure 6 shows the results for all of the subjects and the recognition software using all of the input data. The column results are, in order, correct recognition of the individual segmented letters, correct recognition of the words and, lastly, the case of the word not being recognized but the significant letter in the word correctly identified.

One important result of this experiment was that people do not indeed recognize individual letters perfectly. The machine system is not lagging that far behind human recognition of letters in isolation (65.4% for the machine, 75% for the human subjects). Indeed one subject did not recognize the letters as well as the machine did (65.4% for the machine, 61.5% for the subject).

What is also demonstrated is that human readers utilize contextual information to increase their recognition at the word level. The slight improvement shown by the machine at the word level is the result of lexical constraints. It is apparent that the recognition system will have to exploit other sources of word level information if it is to be as capable as its human counterparts.

4 DISCUSSION

This paper suggests some factors which can be used to improve the performance of a machine system. Experiment 1 showed that ambiguity can actually increase as a machine recognition system is trained on new samples of handwriting. The reason for this is that cursive handwriting is inherently ambiguous. Ambiguity is a necessary consequence of the real variability of written letter forms. The word superiority effect is well established using printed text, experiment 2 confirmed its presence in the reading of cursive script. Human readers are nowhere near perfect on recognizing a set of letters. The machine system is worse than humans, although not far behind and better than one of the human subjects. Machine systems should not be expected to solve the problem of script recognition on the basis of pattern recognition alone; there is a point beyond which further development at this level is futile, and effort should be directed towards the implementation of other sources of information.

Future work will concentrate on the integration of syntactic and semantic information into the recognition system. In part, this will involve the amalgamation of work already completed. [Evett et al, 1992]. A preliminary examination of the data produced by the recognition system suggests, firstly, that the use of syntactic information will be effective and, secondly, that in most cases it is possible to employ very broad grammatical classes, such as tense, or the distinction between singular and plural, in order to produce a strong effect. In other words, it does not appear necessary to introduce a sophisticated parser capable of dealing with more complex or subtle grammatical classes. For example, the following output was obtained from the recognition system:

responsibility
responsibilities

It is not uncommon for the recognition system to produce a list of alternatives that contain morphological variants of one word. One reason for this phenomena is that it is often the case in handwritten text for the end of a word to be written less precisely than its beginning. The use of broad syntactic information in cases such as this will be of obvious benefit.

It is also apparent that the use of semantic information will be useful. A second example taken from the recognition system will serve to illustrate the benefit of semantic information:

right
eight
fight
fright

In this particular case, the recognition software has had difficulty in deciphering the initial letter of the target word. Semantic information can be used in this instance to select the more likely of these alternatives in the given context [Rose & Evett, 1992].

Lastly, it is apparent that for human readers some form of interaction between letter recognition, word information, syntactic information and semantic information is vital to recognition. Text recognition is a prime example of a dynamic system using a feedback mechanism. The current system is bottom-up driven. However, as a general rule it is not effective for information to merely pass up a hierarchy of levels and an alternative approach which is to be explored is to allow higher level processes to influence the operation of lower level processes.

5 REFERENCES

Baron, J., & Thurston, I., (1973), 'An analysis of the word superiority effect', *Cognitive Psychology*, 4, 207-288.

Cattell, J.M., (1886), 'The time taken up by cerebral operations', *Mind*, 11, 220-242.

DeZuniga, C.M., Humphreys, G.W., & Evett, L.J., (1991), 'Additive and interactive effects of repetition, degradation and word frequency in the reading of handwriting', in D. Besner & G.W. Humphreys (eds), *Basic Processes in Reading: Visual Word Recognition*, LEA.

Evett, L.J., Wells, C.J., Keenan, F.G., Rose, T.G., & Whitrow, R.J. (1992), 'Using linguistic information to aid handwriting recognition', in S. Impedovo & J.C. Simon (eds), *From Pixels to Features III: Frontiers in Handwriting Recognition*, Elsevier: North-Holland.

Lindsay, P.H., & Norman, D.A., (1977), *Human Information Processing*, San Diego: Harcourt Brace Jovanovich.

McClelland, J.L., St. John, M., & Taraban, R., (1992), 'Sentence comprehension: a parallel distributed processing approach', in L. Burkholer (ed), *Philosophy and the Computer*, Boulder, Colorado: Westview.

Morton, J., (1969), 'Interaction of information in word recognition', *Psychological Review*, 76, 165-178.

Rose, T.G. & Evett, L.J, (1992), 'A large vocabulary semantic analyser for handwriting recognition', *AISB Quarterly*, 80, 34-39.

Suen, C.Y., (1983), 'Handwriting generation, perception and recognition', *Acta Psychologica*, 54, 295-312.

Tappert, C.C., (1982), 'Cursive script recognition by elastic matching', *I.B.M. Research Development*, 26, 765-771.

The Integration of Knowledge Sources for Script Recognition

G.J. Bellaby & L.J. Evett

Proceedings of IWFHR-IV, 4th International Workshop on Frontiers in Handwriting Recognition, *organising Committee Chair, Jhing-Fa Wang, National Cheng Kung University, ROC, pp 305- 312, Dec 7-9 1994, Taiwan.*

Abstract

Script recognition systems require the use of context to disambiguate input fully. There is a limit on performance beyond which further development at the pattern level will not improve performance significantly. This paper looks at the integration of pattern level, word level and meta-word level information to produce an efficient and robust recognition system. The main topic is the application of information at the word level. There are limits to the recognition of isolated letters. Contextual cues at the word level influence performance of human readers. These cues, in the form of gross word level features can be applied and integrated into the recognition process in order to improve machine performance. Experiment 1 explores methods for deriving and integrating word level information into the recognition process. These methods are developed, evaluated and discussed. Preliminary work on the integration of information taken from the meta-word level is described.

1 Introduction.

Because natural cursive script is inherently ambiguous a machine system cannot recognize individual letters beyond a limited degree of accuracy. Ambiguity is a necessary consequence of the real variability of written letter forms. Machine systems should not be expected to solve the problem of script recognition on the basis of pattern recognition alone; there is a point beyond which further development at this level is futile, and effort should be directed towards the implementation of other sources of information [1].

Letters in isolation, i.e. without any supporting contextual information, are not recognized perfectly by human readers [2]. Contextual information provides a significant advantage for the successful recognition even of isolated words. One instance of the importance of context at the word level is the word superiority effect [1, 3,4]. This is the effect noticeable with human readers whereby letters in words are recognized more easily than letters in isolation or letters in non-words. This effect is well documented although no explanation for the effect has been agreed upon [5]. One factor involved in this effect is that of word frequency [6]. Another is the use of contextual information at the word level. Lexical constraints are not sufficient by themselves to produce the same effect. Human readers use gross word level contextual cues as an important aid to their recognition of words (e.g. information concerning word length, word shape etc.). A machine system that does not exploit contextual cues cannot be expected to show the word superiority effect. The application of gross word level features to the recognition process will bring about a significant improvement in machine performance.

Given sufficient knowledge about contextual cues it becomes possible to identify the word without having to engage in any more detailed analysis of its individual constituents [7]. Integration therefore reduces the need for extensive pattern recognition. It will make the pattern recognition stage much less critical and since pattern recognition cannot hope to be wholly accurate, this can only be of benefit. Many authors have proposed the use of word shape information in pattern recognition systems for cursive script. The present paper has identified whole-word contextual cues from the psychological literature on human reading, and investigates integrating their use into a more traditional, fine-grained pattern recognition system based mainly on letter recognition.

The objective is a machine system which will overcome the particular problems associated with unconstrained cursive handwriting and a large lexicon. The integration of the various different kinds of information associated with reading is an important topic in script recognition. It is certain that a machine recognition system must be able to successfully integrate diverse contextual knowledge sources about a text if it is to reach a level of efficiency comparable to human readers.

2 Experiment 1: Gross Feature Extraction

2.1 Introduction

Human readers do not have to recognize individual letters perfectly. Human readers can use gross word level features to derive a general impression of the word, and in some cases to identify it without more detailed examination [7, 8]. Higher- level context can be used in conjunction with gross feature extraction to make assumptions about the word and from this to derive a candidate. More detailed examination of the word can be used to verify the choice.

Nine sources of information have been identified as being effective at the word level: lexical and word frequency information together with word length, first letter, last letter, and the presence or absence of ascenders, descenders, i-dots and j-dots, and lastly t-crosses and f- crosses. Lexical and word frequency information make a large difference to human performance but are not apparent from the physical information derived from pattern recognition. The other seven sources of information are physical features of a word. This selection was made on the basis of research carried out within the field of cognitive psychology [9, 10, 11].

Given exact information about the length of the word, its first letter, its last letter and whether ascenders, descenders, i- dots, j- dots, t- crosses and f-crosses are present or absent in the target word then perfect detection of these seven features alone can lead to the identification of a single word even in a relatively large lexicon. For example, just over 50 percent of the words in a 15k lexicon can be uniquely identified using the criteria of these seven features. A smaller lexicon (2k) was also compiled using only the higher frequency words (words occurring 50 or more times in a million word corpus). Over 80% of the words in this lexicon can be uniquely identified using the seven features. Within a script recognition system, these features cannot be identified 100% accurately. However, it is likely that even when the information is not entirely accurate it can be beneficial. The present experiment investigated the influence of imperfect information about these features on the performance of a cursive script recognition system. It also investigated a method for integrating these sources of information into a script recognition system.

2.2 Method

The data used in this experiment was derived from the handwriting recognition system developed within the Nottingham Trent University [12]. The handwriting recognition system works by matching letter and segmentation patterns to a pattern database. Hence it is in the database that ambiguity is represented. The handwriting recognition system has a pattern recognizer that produces a set of individual characters, together with segmentation information about these characters, and basic information about the segmentation of words. These characters are combined to produce letter strings which are then filtered according to a lexicon to produce a set of word candidates. The resulting list of words is ranked according to a confidence score which has been given by the recognition system to each of the word alternatives.

Gross features are used to derive a new list of alternatives to add to the existing list generated by the pattern recognizer. No attempt to use the recognizer to directly extract the particular features under consideration is employed. Instead, information about the features is derived from the list of word alternatives given by the existing software. The list of alternatives generated by the recognizer is examined and feature information for each of the words is extracted. The mean, median or mode are used to calculate an average value for the feature or to select its most common occurrence. In some cases the calculation is also weighted by the ranking of the words within the initial list. The reason for using a variety of different methods of calculation is to avoid a simplistic replication of the information held in the initial list. For instance, a reasonable indication of the length of the target word can be found by calculating the mean length of the words appearing in the word list, whilst mode is necessary to derive the most likely first letter and in this case weighting the letters according to the position which their source words appear in the list has proven most effective.

A lexicon is searched for words which match the set of features. A set of probable confusions for each of the features is then used. For instance, substitutions for both the first and last character are introduced for letters which are frequently confused with one another. Likewise, alternative word lengths at lower probability levels are also introduced to deal with errors in the determination of this feature. A relatively high degree of imprecision is allowed in the use of some of the features. For example, certain letters (e.g. 'c', 'e', 'o') are allowed to generate many possible confusions and the estimate of word length is allowed an error margin of 2 or even more characters. It is significant that a fine degree of accuracy is not necessary. The reason for this is that in a great many cases even wide variance will only generate a low number of alternatives. The list of words is allowed to grow in size until a pre-determined threshold is reached. If a word is present in the initial list then it is ignored.

For example, the word 'vainly' was presented to the recognition software and it generated the following list: valley varies variety illness visual It was possible to determine from this list that the target word probably began with 'v', ended with 'y', contained 6 characters, that t-crosses and f-crosses were absent, but that one or more ascender, descender and i-dot or j-dot were present.

This particular set of features, together with their most common confusions, produced the following list: vainly visibly policy plainly poorly purely In this particular case, the target word was actually the first word generated.

The experiment consisted of reasonably clear lower case cursive handwriting being presented to the pattern recognition software. A 15k lexicon based on word frequency and containing every morphological variant of each of the words was used. A sample of writing from 18 subjects was used, each of whom wrote down the same 200 words. The data was written on a Wacom tablet which was under the control of an ink collection program being run on an IBM 486 PC. The data was stored for later processing by recognition software running on a Sun 10 workstation. The samples were of normal, clear cursive handwriting. The pattern recognition software had been trained on the handwriting of just over half of these 18 subjects. Only those cases where the recognition software completely failed to give the target word as an alternative were examined. The feature information was derived from the lists of words output by the recogniser. The words generated by using the gross feature information were then ranked simply according to the order in which they were produced.

2.3 Results

Table 1 shows the percent correct recognition after applying feature extraction in those cases where the pattern recognition software had failed to identify the target. The column results show correct identification of the target by feature extraction alone. The results show a major improvement from a complete failure to provide the target word as an alternative to producing the target word as an alternative in 50.1% of the cases. There has been a significant change from a 100% error rate to a 49.9% error rate. feature extraction top ranked 8.4 top 5 26.3 anywhere in list 50.1

	feature extraction
top ranked	8.4
top 5	26.3
anywhere in list	50.1

Table 1: Percent correct recognition after applying feature extraction

The potential for developing this line of approach is demonstrated by the fact that it became possible to reduce the error rate on the writing sample of one subject to 1 percent.

2.4 Discussion

These results are derived from word lists generated by the pattern recognition software. Even though the recognition software has failed to select the target word as the most likely candidate, it has been possible to use the output of the recognizer to successfully derive the target word in a significant number of cases. This experiment demonstrates therefore how useful gross feature information can be within the recognition process and indicates how different sources of information can be combined to good effect at the word level. Clearly, a more complex ranking procedure could lead to even better results.

The pattern recognition system developed at the Nottingham Trent University is performing at a level comparable to that of human readers in the recognition of letters in isolation [1]. The exploitation of gross word level information could make its performance on the recognition of whole words approach that of its human counterparts.

Two different approaches towards the recognition of cursive handwriting can be identified. In the first approach the recognition process is driven by a pattern recognizer. The main aim of this approach is the identification of what has been written. This aim means that the system is evaluated on the basis of whether or not it has produced the target word. Evaluation is primarily on the basis of one word output, and it is only of secondary importance that the system may give a list of alternatives in which the target word appears. Contextual information is minimal, in that only lexical filtering is carried out. This is a bottom-up approach.

In contrast, the second approach uses the lexicon and applies contextual cues to select words from it. The method of evaluation in this approach is therefore whether the pattern recognition system gave the target word as an alternative and subsequently whether or not contextual cues make it possible to select the target word. This is a top-down approach. The reason why this second method is desirable is that some words are not well written and, in such cases, the second approach will find the word whilst the first method will not.

It is possible to identify a word using gross features regardless of what actual method of recognition is used. The recognition method is irrelevant, except in terms of its accuracy, to the employment of gross features to identify a word. It is not just a case of adding these features to a machine system, indeed some systems may already utilize one or more of the features indicated. Rather these features can be used either alone, or in conjunction with other contextual information, to successfully identify many words. The use of gross feature information is not a mechanical addition to an existing recognition software. These features have, to some extent, been derived from the psychological literature and it has been demonstrated here that they can be effective without recourse to further pattern recognition.

It is possible to direct further examination of the word on the basis of information already derived. This more detailed analysis of the word can refine the features already used, e.g. to count the exact number of ascenders in the word rather than just whether ascenders are present or absent. Alternatively, it is possible to extract further letters or letter sequences from the word. For example, two letters which are frequently confused are 'y' and 'g'. However, when the letter 'g' is placed at the end of a word it typically occurs as part of the letter sequence 'ing' but in the 15k lexicon never as 'lg', whereas 'y' is most commonly preceded by the letter 'l' but only very infrequently as 'iny'. Similar differentiating character sequences can be shown for the other commonly confused letters. Directing a search through the word in this manner has obvious benefits.

Furthermore, a significant number of cases where the seven- feature pattern does not uniquely identify a word are the result of alternative spellings of the same word, e.g. recognize/recognise or utilize/utilise. It is possible to direct attention in these cases to the specific letter (or occasionally letters) which cause these alternative spellings.

Although word level features are currently being used to supplement the operation of the pattern recognition system, it should be realised that the success of applying gross level features to generate a new list of candidate word supports the argument that it is possible to use gross feature detection as the major method of word recognition.

3 Experiment 2: Higher level context

3.1 Introduction

Higher level contextual information can be used to further augment performance. Methods for implementing syntactic and semantic information to aid script recognition have been reported [13, 14]. The present experiment was carried out as a preliminary investigation into the integration of such information into a script recognition system using the current approach.

3.2 Method

Contextual information was used to re-order the list of alternatives: firstly on the basis of broad syntactic classes, and secondly on the basis of syntactic classes together with word frequency. In this first case those word alternatives which were in the same grammatical class or classes as the target word were simply placed above those which were dissimilar. In the second case, words of the same class were also re-ordered according to their word frequency so that candidates with a higher frequency were placed first.

3.3 Results

	feature extraction	feature extraction plus syntax	feature extraction plus syntax and word frequency
top ranked	8.4	27.7	29.3
top 5	26.3	43.8	46.0
anywhere in list	50.1	50.1	50.1

Table 2: Percent correct recognition after applying feature extraction both with and without context

The column results in table 2 show, in sequence, correct identification after the lists of alternatives generated by feature extraction are sorted by syntactic class, and lastly after the lists are sorted by syntactic class and word frequency. These results demonstrate that syntactic information and word frequency information can be used successfully to select the target word from a list of alternatives. The use of both of these sources of information in conjunction place the target word at the top of the list of alternatives in 29.3% of the cases, and in the top 5 in 46% of the cases.

3.4 Discussion

Further work is underway to improve the way in which gross feature information is derived. As work continues the results shown in Table 2 will be expanded to include the effects of introducing semantic information, both on its own and in conjunction with syntactic information and word frequency. Syntax has only been applied in a very simplistic manner here. Markovian approaches to implementing syntactic information are such that the probability of syntactic class for any word position can be estimated on the basis of its word context [13]. Experiments are underway in order to integrate this process into the system. Evaluation of the use of probabilistic syntactic information will be carried out. Probabilistic semantic information can also be applied in this manner [14].

4 General discussion

This paper suggests some ways in which the performance of a machine system can be improved. Experiment 1 demonstrates that it is possible to use a set of gross word level features to identify the target word even though the pattern recognition software has completely failed to give the target word as an alternative. It has proven possible to derive useful information about the gross features of a target word from the list of alternatives suggested by the pattern recognizer even though the recognizer did not identify the target word itself. The feature based method therefore produces additional candidates to the pattern recognition method, even though it is based on information derived from the pattern recognizer. In a significant number of cases this gross feature information can be used to derive the target word. This experiment also shows that integrating word level information into the recognition process is effective.

Experiment 2 looked at the integration of word frequency and syntactic information into the recognition system. A preliminary examination of the data produced by the recognition system suggests, firstly, that the use of syntactic information will be effective and, secondly, that in most cases it is possible to employ very broad grammatical classes, such as tense, or the distinction between singular and plural, in order to produce a strong effect.

Future work will continue the process of integrating contextual information into the recognition process. In part, this will involve the amalgamation of work already completed. [15], and it will be a continuation of the approach outlined in Experiment 2.

Lastly, it is apparent that for human readers some form of interaction between letter recognition, word information, syntactic information and semantic information is vital to recognition. Script recognition is a prime example of a dynamic system using a feedback mechanism. As a general rule it is not effective for information to merely pass up a hierarchy of levels and an alternative approach which the present work is pursuing, is to allow higher level processes to influence the operation of lower level processes, and contribute information to the recognition process.

5 References

1. Evett, L.J., & Bellaby, (1994), 'The integration of knowledge sources for word recognition', *IEE European Workshop on Handwriting Analysis and Recognition: A European Perspective.*

2. Suen, C.Y., (1983), 'Handwriting generation, perception and recognition', *Acta Psychologica*, 54, 295-312.

3. Cattell, J.M., (1886), 'The time taken up by cerebral operations', *Mind*, 11, 220-242.

4. Baron, J., & Thurston, I., (1973), 'An analysis of the word superiority effect', *Cognitive Psychology*, 4, 207-288.

5. Monsell, S., (1991), 'The nature and locus of word fequency effects in reading', in D. Besner & G.W. Humphreys (eds.), *Basic Processes in Reading: Visual Word Recognition*, LEA.

6. Broadbent, D.E., (1967), 'Word frequency effect and response bias', *Psychological Review*, 43, 1-15.

7. Rayner, K., Well, A.D., Pollatsek, A., & Bertera, J.H., (1982), 'The availability of useful information to the right of fixation in reading', *Perception & Psychophysics*, 31, 537-550.

8. O'Regan, J.K., (1979), 'Eye guidance in reading: Evidence for the linguistic control hypothesis', *Perception & Psychophysics*, 25, 501- 509.

9. DeZuniga, C.M., Humphreys, G.W., & Evett, L.J., (1991), 'Additive and interactive effects of repetition, degradation and word frequency in the reading of handwriting', in D. Besner & G.W. Humphreys (eds), *Basic Processes in Reading: Visual Word Recognition*, LEA.

10. Lindsay, P.H., & Norman, D.A., (1977), *Human Information Processing*, San Diego: Harcourt Brace Jovanovich.

11. Morton, J., (1969), 'Interaction of information in word recognition', *Psychological Review*, 76, 165-178.

12. Powalka, R.K., Sherkat, N., Evett, L.J., & Whitrow, R.J., (1993), 'Dynamic cursive script recognition: A hybrid approach', *Sixth International Conference on Handwriting and Drawing*, Paris, France, July.

13. Keenan, F. & Evett, L.J., 'Applying syntactic information to text recognition', *Computational Linguistics for Speech and Handwriting Recognition*, April, AISB 1994 Workshop Series.

14. Rose, T.G. & Evett, L.J, (1993), 'Semantic analysis for large vocabulary cursive script recognition', *Proceedings of the Second International Conference on Document Analysis and Recognition*, Tsukuba Science City, Japan

15. Evett, L.J., Wells, C.J., Keenan, F.G., Rose, T.G., & Whitrow, R.J. (1992), 'Using linguistic information to aid handwriting recognition', in S. Impedovo & J.C. Simon (eds), *From Pixels to Features III: Frontiers in Handwriting Recognition*, Elsevier: North- Holland.

Multiple knowledge sources for word recognition

G.J. Bellaby, L.J. Evett & R.K. Powalka

Proceedings of IEE Workshop on Handwriting Analysis and Recognition- A European Perspective, May 1996, organised by Professional Group E4 (Image Processing and Vision) pp 9/1-9/3.

Traditional character based handwriting recognition systems are geared towards giving the target word as the top ranked choice in a set of likely candidates. This is, of course, reasonable since output from a handwriting recognition system has to be judged on the basis of the target word being top ranked and this calls for one forced choice. However, there is a disadvantage to this approach since the effort to place the target word at the top rank comes at a price. That is that the word lists generated by traditional handwriting recognition systems are not well suited for post-processing (e.g. syntactic and semantic analysis). For a system to be suitable for post-processing, it is important that the target word is found, regardless of its rank. A traditional handwriting recognition system can be considered to be discriminatory (target top ranked) but fragile: it tends either to get the target word correct (to place the target word at the top of the list of alternatives), or to fail to identify the target word at all.

It is not possible to get complete disambiguation of handwriting from pattern recognition alone since a written word can be interpreted in a number of different ways. The combination of several sources of information, each of which is capable of extracting a different characteristic of cursive handwriting, is more likely to be successful than pattern recognition alone. The way to produce a machine system which is both discriminatory and robust (target found even when it is not top ranked) is to combine different, but complementary recognition methods. It is only by integrating different sources of information that a stronger, more robust, machine system can be developed. Three sources of information are considered in the present paper: character- segmentation information, word shape information and lexical information. The methods used to extract these three sources of information are, respectively, a traditional pattern recognizer, a whole word recognizer and a method which uses word level contextual cues.

The pattern recognizer applies an interactive method which combines segmentation, letter recognition and lexical look-up processes. Powalka calls this recognition method "multiple interactive segmentation" [Powalka, et. al., 1993; Powalka, 1995]. The pattern recognizer generates sets of characters together with segmentation information about these characters. Letter patterns are then matched against a database of known patterns. In this manner, letter sequences are built up. A lexicon is used to verify the letter sequences and only known letter combinations are processed further [Wells, et. al., 1990]. The final outcome of the pattern recognizer is a list of word alternatives, each of which has an associated confidence score. The pattern recognition system is highly discriminatory: in those cases where it has recognized the intended word it tends to place the target word at the top of the ranked list of word alternatives. The pattern recognition system orders word candidates solely on the basis of their physical characteristics.

A whole word, or holistic, recognizer has also been created. The holistic recognizer exploits word shape information [Powalka, et. al., 1994; Powalka, 1995]. This includes zoning information, which is used as a guide for locating ascenders and descenders, attempts to estimate word length, and the application of independent letter verification procedures. The holistic recognizer uses the physical characteristics of the input but bypasses the exacting requirement of identifying all of the characters of a word. Instead, it favours recognizing the overall shape of the word and subsequently attempting to verify individual characters in order to produce a 'best fit' of word to shape.

A word level method (WLM) has also been developed. In part, the WLM uses information which is also available to the pattern recognizer, but it re-organizes this information and structures it in a different way. It also applies it in a different fashion to that of the pattern recognizer. A number of cues which are useful at the word level have been identified: word length, first letter, last letter, and the presence or absence of ascenders and descenders, i-dots and j-dots, and t-crosses and f-crosses. The WLM uses lexical information as well as information from the pattern recognizer. A method to derive values for these word level cues using the list of candidates has been developed. The WLM can therefore be implemented easily using any existing pattern recognizer which generates a list of alternatives. An alternative source for these values is direct pattern recognition.

Word level cues are used in conjunction with, firstly, knowledge of the kind of confusions generated by the recognizer and, secondly, knowledge of the kind of word patterns which are present in a lexicon (just over 20% of the words in a 15,000 word lexicon can be uniquely identified using the criteria of these seven cues. The average number of words selected using the cues is 2.59). Word level contextual cues are used in three ways. Firstly, they are used to re-order the list of word alternatives generated by the pattern recognizer. Secondly, word level contextual cues are used to search the lexicon and so generate new candidates to add to the existing ones created by the pattern recognizer. The lexicon is then searched using these cues. Viable candidates can be derived, even when the recognizer did not identify the target word. Lastly, the pattern recognizer can experience catastrophic failures where it completely fails to generate any output. A candidate list is generated using word level contextual cues in the same way as before. The WLM tends to generate the intended word, but requires additional support (e.g. word frequency information) in order to increase the probability that the intended word appears towards the top of the list.

	traditional	holistic	WLM	traditional + holistic	traditonal + WLM
top rank	61%	43%	24%	63%	64%
total (100)	72%	79%	74%	88%	90%

The accuracy of the three sources of information used by the machine system differ. However, different kinds of accuracy have been identified. The pattern recognizer can be considered to be highly discriminatory because it can often specifically identify the target word; it tends to place the target word at the top of the ranked list of word candidates. However, the pattern recognizer is fragile because it often fails to recognize the target word. The holistic recognizer is less fragile, but lacks the fine-grained recognition abilities of the pattern recognizer. The WLM is more robust, but less discriminatory, than the pattern recognizer; it tends to find the intended word but has problems ensuring that the intended word appears towards the top of the list. The table shows recognition performance for each recognizer alone, and some combinations. It is worth noting that, in the case of the combined traditional and WLM method, additional processing is possible which further improves performance [Bellaby, et. al., 1996].

Different approaches towards the recognition of cursive handwriting have been identified. In the first approach the recognition process is driven by a traditional pattern recognizer. This is a bottom-up approach. The second approach relaxes some of the constraints of character and segmentation recognition in favour of word shape recognition and subsequent character verification. The third approach uses the lexicon and applies contextual cues to select from this list of words. This is a top-down approach. The reason why the latter two methods are desirable is that some words are not well written and, in such cases, they will find the word whilst the first method will not. The combination of different sources of information also offers the opportunity to develop interactive processes within the system,

It has been demonstrated that it is possible to successfully merge the outputs of the pattern recognizer, the holistic recognizer and the WLM. The holistic recognizer and the WLM are not intended to be alternatives to the pattern recognizer. A specific point in the recognition process for the application of these approaches has been identified. It is argued that pattern, whole word and word level recognition play distinct but complementary roles in the machine recognition process. The consequence of integrating these different approaches is an increase in coverage without any loss in precision. In other words, a handwriting recognition system which is both discriminatory and robust has been created. Recognition methods which display significantly different characteristics can be integrated to improve machine performance.

References

G.J. Bellaby, L.J. Evett and R.K. Powalka, (1996), `Coping with ambiguity and error in script recognition', *5th International Workshop on Frontiers in Handwriting Recognition, University of Essex, England* (forthcoming).

R.K. Powalka, (1995), `Algorithm toolbox for on-line cursive script', Unpublished PhD Thesis, Nottingham Trent University.

R.K. Powalka, N. Sherkat, L.J. Evett & R.J. Whitrow, (1993), 'Multiple word segmentation with interactive look-up for cursive script recognition', *Second*

International Conference on Document Analysis and Recognition ICDAR'93, Tsukuba Science City, Japan, October, 196-199.

R.K. Powalka, N. Sherkat & R.J. Whitrow, (1994), 'Feature extraction: on the importance of zoning information in cursive script recognition', In S. Impedovo (ed.), *Progress in Image Analysis and Processing III, World Scientific,* 342-349.

C.J. Wells, L.J. Evett, P.E. Whitby & R.J. Whitrow, (1990), 'Fast dictionary look-up for contextual word recognition', *Pattern Recognition*, vol. 23, no. 5, 501-508.

Coping with ambiguity and error in script recognition

G.J. Bellaby, L.J. Evett & R.K. Powalka

Proceedings of IWFHR-V, 5th International Workshop on Frontiers in Handwriting Recognition, University of Essex, England, 1996.

Abstract

Cursive handwriting is characterized by character strings with ambiguous boundaries and considerable variation in letter form. A proportion of the output of a pattern recognizer will therefore be ambiguous (target identified, but not selected as the top ranked choice) or incorrect. The approach proposed in this paper can improve upon this set of ambiguous or incorrect results for on-line cursive handwriting. The method uses word level contextual cues in order to construct a new list of word candidates. This word level method (WLM) uses these cues to search a lexicon in order to generate new candidates. No attempt is made to correct an output word, rather new, probable, alternatives are generated. Viable candidates can be derived in this way, even when the recognizer did not identify the target word. The new candidate list is then merged with the output of the pattern recognizer. Finally a letter verification procedure is applied to the resulting merged list. The result of this is that system accuracy and robustness are both improved. This shows that high- and low-level cues can be successfully integrated.

1 INTRODUCTION

A conventional, character-based pattern recognizer is geared towards placing the target word as the top ranked choice in a set of likely candidates. (Details about the pattern recognizer are given elsewhere [Powalka, 1995]). This strategy is, of course, quite reasonable since the main criterion for evaluation must be the recognition of the target word. However, there is a cost to be paid. That is that the word lists produced by the pattern recognizer are not apposite for post-processing (e.g. syntactic and semantic analysis) because further selection is not possible unless the target word occurs within the list of alternatives. Post-processing is necessary because handwriting recognition systems are not yet accurate enough. Post-processing allows further selection to made from a list of word alternatives and so increase levels of accuracy.

The pattern recognizer is discriminatory since it tends to choose the target word as the top ranked choice. However, the pattern recognizer is not always successful because of the ambiguity of cursive handwriting. The pattern recognizer tends either to place the target word at the top of the list of alternatives, or to fail to identify the target word at all. It is only in a relatively small amount of cases that the pattern recognizer identifies the target word but places it at a lower rank. The pattern recognizer can therefore be characterized as fragile. The WLM displays different characteristics from that of the pattern recognizer. The WLM lacks the fine-grained recognition abilities of the pattern recognizer, i.e. it is not discriminatory. The way to create a machine system which is both discriminatory and robust is to combine different, but complementary, recognition methods.

2 THE WORD LEVEL METHOD

The WLM uses seven cues to derive new word candidates: word length in characters, first letter, last letter and the presence or absence of ascenders, descenders, i-dots and j-dots, and t-crosses and f-crosses. These cues are not physical, or at least not entirely physical, but rather are abstract [cf. Humphreys et. al., 1990]. For example, length is expressed in terms of the number of characters. Word length is not a physical characteristic of a word's length, shape, or the ratio of height to width, but an abstraction. There is no simple relation to the actual physical length of the pattern, but an abstract representation of length based on the number of characters identified in a word. The two cues first and last letter are also not items of physical information; they are characters, identities rather than physical patterns. The last four cues do have a physical component. However, they are also abstract cues which preserve the sort of information retained across letter confusions, e.g. the fact that tall letters tend to be confused with other tall letters [Bouma, 1971]. Therefore, these cues are not entirely physical. For instance, they are position independent.

A pragmatic argument, rather than a strong theoretical one, has been used to select this exact set of cues. The lexicon was examined to see which cues could be used to partition the lexicon efficiently without simply replicating the information used by the pattern recognizer. These seven cues are surprisingly effective. Over 20% of the words in a 15,000 word lexicon can be uniquely identified using just the criteria of these seven cues and over 50% of the words are in groups which have 4 members or less. The largest group of words delineated by the cues has 35 members, and only one group exists at this size. The average number of words delineated by the cues is 2.6.

Two sources exist for these word level cues. A method for deriving word level cues from the list of candidates generated by the pattern recognizer has been developed. The derivation of word level cues from a candidate list is a method which can be implemented without the need for a new pattern recognizer. The WLM can therefore be implemented easily using any existing pattern recognizer which generates a list of alternatives. The information which is used by the pattern recognizer can also be used by the WLM. However, the information is being used in two completely different ways. The WLM re-organizes this information for its own purposes and structures this information in a different fashion to that of the pattern recognizer.

An alternate source for the word level cues is direct pattern recognition. It has proved possible to use relevant data from a pattern recognizer to derive word level cues. For example, the word length is calculated from the number of approximately vertical strokes in a word which are directed downwards. The average number of letters per vertical bar can be calculated. For example the letter `m' is typically written using 3 vertical bars, `l' with 1 vertical bar, and `k' using 2 vertical bars.

The WLM has three ways of applying word level cues. Firstly, the WLM uses word level cues to re-order the original word list generated by the pattern recognizer. Secondly, word level cues are used to generate new candidates which are subsequently merged with the re-ordered word list. Thirdly, the pattern recognizer can experience catastrophic failures where it completely fails to generate any output. Word level contextual cues are used to generate a candidate list which is used instead of the pattern recognizer.

The seven cues are used in conjunction with lexical information, e.g. the target begins with `c', ends with `t', contains three characters, contains an ascender and a cross, but does not contain a descender or a dot, i.e. "cat", "cot", "cut". A set of probable confusions for each of the cues is used based on the known accuracy of detection and, in the case of length, first and last, likely confusions. For instance, alternative word lengths at lower probability levels are introduced to deal with errors in the determination of this cue. A confusion matrix is used for both first and last characters reflecting the letters which are frequently confused with one another. Likewise, a relatively high degree of imprecision is allowed in the use of some of the cues. For example, certain letters (e.g. `c', `e', `o') are allowed to generate many possible confusions. Alternative patterns are then generated using these probabilities. Some of the generated patterns can be rejected, e.g. in those cases where one or more of cues are contradictory, e.g. "contains the letter `d'" and "does not contain an ascender", or where the generated pattern does not occur in the lexicon. Finally, the patterns are sorted according to their probability and a lexicon is searched for the words which match each of the generated patterns. In this manner a list of alternatives was produced for each of the words under examination.

3 EXPERIMENT 1: COMBINING THE METHODS

An experiment was conducted where lower case cursive handwriting was presented to pattern recognition software. A 15,000 word lexicon based on word frequency and containing every morphological variant of each of the words was used. The maximum possible number of words in the list of alternatives generated by the pattern recognizer was limited to 100. A sample of writing from 18 subjects was used, each of whom wrote down the same 200 words (a total of 3,600 samples). The legibility of the data ranged from neat to poor handwriting (target words top ranked by the pattern recognizer ranged from 92% for the best writer, down to 19% for the worst writer). The data also represented a range of writing styles.

Word level contextual cues for each of the targets were derived. From this a set of alternative word candidates was generated. A confidence score was given to each of the alternatives generated by the WLM based on the probability of the pattern's occurrence and the word frequency of the alternative. The list of words generated by the pattern recognizer was re-ordered using the criteria of the word level cues and subsequently merged with the alternatives produced by the WLM using a simple voting method. That is, the resulting merged list was ordered on the basis of the various confidence scores given to each word alternative. Finally, the merged word lists were presented to a further recognizer which applied letter verification procedures on the handwriting samples. Candidates were increased in rank on the basis of the proportion of their characters which the letter verifier recognized within the sample.

The letter verification methods used in this experiment were developed for a holistic, or word shape, recognizer [Powalka, et. al., 1994; Powalka, 1995]. Letter alternatives are located and recognized by the recognizer. The confidence scores of the located letters are combined with the scores obtained from merging the WLM and the pattern recognizer. Merged score and scores obtained from letter verification are averaged to produce a final candidate score. In this way letters from both sources are boosted with respect to letters proposed by only one source.

4 RESULTS

The following table shows the percent correct recognition for the individual methods, a combination of the two, and a combination of the plus letter verification.

	pattern recognizer	WLM	pattern recognizer + WLM	letter verification
top rank	61%	24%	64%	66%
top 10	71%	55%	80%	84%
top 100	72%	74%	90%	90%

Table 1: Recognition results for the methods on their own and in combination

Results for merging the pattern recognizer and the WLM show an increase in all three columns. Letter verification caused a further increase in the proportion of target words top ranked and the proportion of target words in the top 10. Since ceiling effects will minimize any possible improvements, these figures are significant. Just considering those cases where the recognizer failed to produce the correct word, the WLM derives the word 51% of the time. Other methods of merging the lists could further increase the proportion of target words top ranked; the use of other knowledge sources could also bring improvement. The fact that the WLM increased the proportion of targets found increases the scope for such improvement.

5 DISCUSSION

The WLM cannot be as discriminatory as the pattern recognizer since it selects groups of words from the lexicon using the criteria of the seven chosen cues. The WLM is not more robust than the pattern recognizer on its own (considering only the top 10 alternatives, although it is when the top 100 candidates are taken into consideration). However, the WLM is a very different source of information from the pattern recognizer and it can therefore be effectively combined with the pattern recognizer to create a system which is robust. However, the results for the WLM on its own do not include the re-ordering of the original list by the WLM, just the generation of word candidates. The results for the combination of the pattern recognizer and the WLM do include re-ordering.

This experiment demonstrates that integrating word level information into the recognition process can be effective. The pattern recognition system orders word candidates solely on the basis of their physical characteristics. The pattern recognition system can be considered to be a highly discriminatory method; in those cases where it has recognized the intended word it tends to place the target word at the top of the ranked list of word alternatives. The WLM is, by contrast, less discriminatory than the pattern recognizer but can cope with ill- formed handwriting, e.g. one poorly written character can cause the recognizer to misidentify a word, but will not necessarily cause the WLM to fail. The WLM can also deal with poorly delineated characters, i.e. it bypasses some of the problems involved in segmentation. However, the WLM cannot hope to compete with the pattern recognizer in its ability to place the target word at, or near the top of, the ranked list of word alternatives because the WLM does not have the selectiveness of the pattern recognizer. The merging of the list of alternatives generated by the WLM with the list of alternatives generated by the pattern recognizer leads to an increase in robustness (i.e. finding the intended word) and an increase in the proportion of targets words top ranked. Whilst a bias effect towards high frequency words has been introduced, it is restricted to confusions. In other words, physical cues prevail over frequency factors.

Attempts have been made previously to develop a two-stage recognition system [e.g., Hull, et. al., 1983]. Additions to the recognizer developed at the Nottingham Trent University, such zoning information and other types of feature detection, also implement certain types of word level feature extraction. However, the use of word level information is not a mechanical addition to an existing recognition software. It is not envisioned, for example, that additions to Nottingham Trent's pattern recognition system will, or can, supersede the WLM. It has been demonstrated here that these cues can be effective without recourse to further pattern recognition. They cannot be derived through pattern recognition alone.

6 CONCLUSIONS

It has been demonstrated that it is possible to use a set of word level cues to identify target words even though the pattern recognition software has completely failed to give the target word as an alternative. Syntactic and semantic information can be used to make a selection from a list of alternatives. However, this is only the case if the target word occurs as one of the alternatives. The reduction in the proportion of target words unrecognized caused by the WLM is therefore of considerable benefit to the machine system. The accuracy of the various sources of information used by the machine system differ. The pattern recognizer can be considered to be highly discriminatory because it can often specifically identify the target word; it tends to place the target word at the top of the ranked list of word candidates. However, there are relatively few cases in which the pattern recognizer identifies the target word but places it at a lower rank. The WLM is more robust, but less discriminatory, than the pattern recognizer; it tends to find the intended word but has problems ensuring that the intended word appears towards the top of the list. It has been demonstrated that it is possible to merge successfully the outputs of the pattern recognizer and the WLM. The consequence of integrating these two methods and the use of letter verification is an increase in overall accuracy and robustness. In other words, the proportion of target words found is increased and the proportion of target words top ranked is also increased. Two recognition methods which display significantly different characteristics can be integrated to improve machine performance. It is therefore argued that the WLM is a viable way to cope with ambiguous or incorrect output from a pattern recognizer.

7 REFERENCES

H. Bouma, (1971), `Visual recognition of isolated lower-case letters', *Vision Research*, vol. 11, 459-474.

G.W. Humphreys, L. J. Evett & P.T. Quinlan, (1990), `Orthographic processing in visual word identification', *Cognitive Psychology*, vol. 22, 517-560.

J.J. Hull, S.N. Srihari & R. Choudhari, (1983), `An integrated algorithm for text recognition: comparison with a cascaded algorithm', *IEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-5, no. 4, July 1983, 384-395.

R.K. Powalka, (1995), `Algorithm toolbox for on-line cursive script', Unpublished PhD Thesis, Nottingham Trent University.

R.K. Powalka, N. Sherkat & R.J. Whitrow, (1994), `Feature extraction: on the importance of zoning information in cursive script recognition', In S. Impedovo (ed.), *Progress in Image Analysis and Processing III*, World Scientific, 342-349.