A Computational Model of the First Stage of Learning to Read

Stephen Cassidy, Peter Andreae
Department of Computer Science

G. Brian Thompson
Department of Education
Victoria University of Wellington

Address for Correspondence:
Stephen Cassidy,
Speech Hearing and Language Research Centre,
Macquarie University,
North Ryde, NSW 2109,
Australia
Email: steve@srsuna.shlrc.mq.edu.au
Phone: +612 8508729
Fax: +612 8508240

Abstract

Current theories of reading development leave unspecified many of the details of the procedures used by children as they learn to read. This lack of detail prevents many questions being answered about the relationship between various sources of knowledge and the child's new reading skills. One way of forcing details to be considered is to build a computational model; this paper describes such a model of the very first part of a child's reading development, corresponding perhaps only to a few weeks of developmental time. One aim of this work is to show how children apply the skills they have before they start to read to their very first experiences with print.

The model implements a visual word recognition procedure based on a lexicon of stored representations accessed via visual cues. Words are stored initially as an unordered set of letter tokens. This representation is incomplete in that some letters may be replaced by markers and others may be omitted altogether. As the reading procedure develops, the representation becomes more accurate and the order of letter tokens is also stored. The way that words are selected from the lexicon changes so that initial and final letters are used as cues. This developmental pattern is explored in a series of 'snapshot' simulations which model the procedure under a given set of parametric assumptions. The simulations are used to predict the characteristics of the reading procedure, including the types of errors made at each stage. The error profile of the hypothesised developmental sequence for visual reading is shown to correspond to published data from longitudinal studies of children's reading. Finally, an account of development during this first stage of reading is presented in terms of Karmiloff-Smith's representational redescription framework.

Introduction

When children begin to read they know nothing of the orthographic structure of their language. In learning to read they not only have to learn how words appear on the page, but also the best way to store and access these words. They may gain some clues from their knowledge of spoken words or alternatively from the way that they recognise pictures or objects. The problem for the reader is applying these ideas to the new task and then making their solution fit the problem as well as possible. A real, optimal solution requires fore-knowledge of orthographic structure. Our beginning reader does not have this and so must make do with whatever skills can be applied to the new task.

Some theories of reading development have settled on the idea of developmental stages. So Frith (1985) suggests that children begin to read using distinctive graphic features, which may or may not be letters, and that letter order is largely ignored. Frith calls this reading using Logographic skills. At the subsequent stage, the reader uses skills of phonological mediation, recoding letters into phonemes, called Alphabetic skills. Lastly, Orthographic skills are acquired. These relate stored letter sequences of the word directly to morphemes, without the mediation of phonology. In all of this work however, questions remain about the details of the skills used at each stage, and about the relationships between the stages. Is it the case, for instance, that the Logographic stage is mandatory or can a suitably skilled reader begin reading in the Alphabetic stage? If the Logographic stage is mandatory, what skills or knowledge does it provide that are so important to the development of later reading skills?

Stuart and Coltheart (1988) addressed the issue of a mandatory Logographic stage in a longitudinal study of a number of beginning readers. They found some readers who appeared to be using a phonologically mediated strategy very early in their development. These readers were capable of manipulating sub-word phonology and occaisionally made non-word responses to word targets. The reading errors that these children made tended to share first or first and last letters with the target word, and the proportion of this kind of error increased as the child developed. Stuart and Coltheart suggested that these errors were phonologically derived and, since some readers in their study began reading in this way, that this phonologically mediated procedure did not depend on a preceding logographic procedure. The suggested mechanism involved the child making predictions about the orthographic form of a word based on knowledge of the word's pronunciation and some spelling (sound to letter) knowledge. These expectations would tend to be accurate at least in the first and last letters as these sounds are the ones most easily segmented by a young child (Stanovich, Cunningham, & Cramer, 1984). Now, when the child sees the word for the first time in print, the expected orthographic form can be used to recognise the word.

If we examine this proposal in mechanistic detail we find some flaws even without considering any additional behavioural evidence. Notice, that this model assumes that a mechanism already exists for recognising words based on these orthographic expectations. This mechanism knows how to store orthographic representations of words and match them against a stimulus. It knows that letters are the important units and that they are left-to-right ordered in written words. In short, this mechanism is quite an accomplished word recognition system in its own right. It seems unlikely that such a mechanism could exist without it having been used for reading some familiar words prior to the insertion of the expectations suggested by Stuart and Coltheart. If it is used, then the child has *begun* to read using visual orthographic representations, not by phonological mediation as suggested by Stuart and Coltheart. It may be that the child doesn't take long to develop these skills, but in an account of reading development which seeks to account for the source of children's knowledge, this visual orthographic stage cannot be overlooked.

This point is moot in this particular case as Stuart later found evidence against the expectation building mechanism (Stuart, 1990) but it serves to illustrate that it is difficult to propose a phonological reading procedure as a true first strategy for a child. A realistic procedure requires a reasonable level of orthographic knowledge to work; a level of knowledge that cannot be assumed in a non-reading child.

The aim of our work is to explore the process of reading development with the specific goal of providing an account which allows a non-reading child to develop a level of orthographic knowledge sufficient to provide a basis for further development. Our account will attempt to make explicit the knowledge and skills that a pre-reader has that enable reading development to begin. We hope to provide an account of the knowledge gained in the first stage and how it might affect subsequent development. Our methodology is to build a series of computational models to explore parts of the theory; this paper reports on the first of these models: a model of the initial word recognition procedure.

What Children Know Before They Start to Read

It is hard, if not impossible, to pinpoint the exact time when a child starts being a reader. The beginning of reading is clearly a gradual transition from being unable to recognise any printed form as a word to 'reading' environmental words (eg. McDonalds) to reading stories from books at school. There would seem to be no point at which the child suddenly has the requisite knowledge and skills for reading: wherever the line is drawn the child has at least some knowledge relevant to the 'new' task. However, a starting point must be chosen: this project is concerned with the development of the child from the time when words are seen as being composed of letters and begin to be associated with word meanings and pronunciations in a systematic way. This section discusses the knowledge that can be assumed for a child at this stage of development.

On the surface, early word recognition seems to be similar to general object recognition: names are associated with graphic forms, possibly consisting of separately identifiable sub-parts. The two tasks become differentiated only when the special status of letters is learned. When they start to read, children are already accomplished object and picture recognisers.It is reasonable to assume then that the child is able to distinguish a number of the letter forms from one another, although by no means all. Since the child does not know much about the nature of orthography, it is unlikely that different allographs (alternative forms of a letter or grapheme, for example, upper and lower case) will be recognised as being equivalent. So for instance, *THE* and *the* will be seen as distinct items which share the same pronunciation in the same way as homophones like *rose* and *rows*. Eventually, the child will learn to categorise written glyphs as letters independent of allographic variation. Another feature of written English, the left-to-right order of letters encoding sounds, must also be learned from experience with print or through direct teaching.

Before they are capable of any sort of reading task, children are able to recognise a sizeable vocabulary of spoken words (around 5000 at age five according to Aitchison (1986, p139)). This implies that the child knows, at least implicitly, that a word is a unit of the language and so can begin to associate written forms with their spoken counterparts. It might be further assumed that the child knows about sub-word units such as syllables or phonemes and so might transfer these concepts to sub-word units in written words; that is the child might expect that written words will be made up of parts just like spoken words are. However, it is not at all clear that a pre-reader has such well developed sub-word phonological knowledge. Some researchers argue that experience with words in print is a prerequisite for such knowledge (Morais, Cary, Alegria, & Bertelson, 1979), while others suggest that children must be able to manipulate sub-word phonology before they can begin to read (Bradley & Bryant, 1985). Liberman et al. (1974) found that pre-readers are able to segment pronunciations into syllables and phonemes but are not very good at it: they get better after they learn to read. The model to be presented here does not assume anything about the state of the child's phonological knowledge before they start to read. The relation of phonological knowledge with the developing reading procedure will be discussed later.

Children's Initial Representations of Print Words

In English orthography, a word token is written as a sequence of letters, ordered from left to right. In order to recognise a token as an English word, the reader must have a stored representation of that word. This representation is a record of the distinguishing features of the word stored in such a way as to enable the reader to differentiate that word from all other letter strings. The representation need not record everything that could be said about a print word — colour, for instance, is unlikely to be of importance. The representation need not take the form of a declarative description of the word; it could, for instance, be stored as a procedure which outputs a particular symbol when the print word is given as input. Such a procedure could be implemented in a number of ways, for example as a traditional symbolic program or as a neural network model. The goal of this section is to set out the external properties that this representation must exhibit; such properties can be inferred from observations on the nature of the reading task and the way that children perform it when they start to read. These properties should guide the implementor of any model of this stage of reading, whatever the chosen implementation paradigm.

Distinctions will be made here between an ideal representation of the print word and the representation which children seem to have. In a mature reader, these are likely to be similar if not the same. Since the child begins by knowing nothing about orthography, it is unlikely that the first representations used will be ideal in this sense.

Perhaps the least controversial aspect of the representation of a print word is that it must record which letters are contained in the word token. This is not as simple as it may seem, however. For example, the word *cabbage* contains the letters *[a,b,c,e,g]* – a complete representation needs to record the fact that there are two distinct *a*'s and two distinct *b*'s: it needs to record the *letter tokens* in the word.

Observations of children's word recognition errors show that at least in the early stages of development, they do not take notice of all of the letters in a word. Errors often share only the first or last letter with the target and occasionally have only a salient feature, such as an ascender or a descender in common (Stuart & Coltheart, 1988; Campbell, 1987). This suggests two interpretations: either the child's stored representation of the word only contains these salient features or the stored representation is complete but the procedure that matches the representation of a stimulus with the stored one is willing to ignore some details[1]. Evidence from children's spelling (Gentry, 1982) suggests that, if children use the same lexical representation for reading and spelling, then it is this representation that lacks detail; for instance, a child might produce *ct* as a spelling for *cat.* This model assumes that the child's stored representation of words is incomplete at an early stage and becomes more complete as reading skills are acquired.

Another aspect of the structure of a print word that should be recorded in a complete representation is the ordering of the letter tokens. Without a representation of ordering, no distinction can be made between words which contain the same letter tokens (for instance, *meal, male* and *lame*). However, children do not always pay attention to letter order or position, as noted by Seymour and Elder:

> "It appeared that at a certain point in development the *position* of salient letters was not used as a discriminatory feature. For example the letter "k" was a salient feature for identification of the word "*black*". In the 16th week of schooling, LBH responded "*black*" to *likes, think* and *thank*, and also to the non-word targets *bkacl, eadhk, pjoek* and *htoek.*" (Seymour & Elder, 1986)

---

[1] This assumes that the stimulus representation has the same properties as the stored representation. If the two do not correspond, then another interpretation is possible: that the stored representation is complete but the stimulus representation is not. If this were the case, then two representation building procedures would be necessary: one for stimuli and one for internal representations. It seems more natural to propose only one such procedure and have the two representations correspond.

Stuart and Coltheart (1988) also observe substitution errors sharing salient letters but ignoring position. In their longitudinal study the proportion of such errors decreases with time while the proportion of errors where the position of the common letters is maintained increases as the children develop. Again, this could be interpreted as a procedure which ignores letter order and position when doing word recognition or a representation where this information is not recorded. In either case, it must be possible for ordering to become important in order to replicate Stuart and Coltheart's observation and in order to account for mature performance which necessarily takes account of this information.

When children begin to read, they seem to know which words they can read and which they can't: when they make an error the response is likely to be one of their reading words, rather than another word from their spoken vocabulary (Campbell, 1987; Seymour & Elder, 1986). This implies that children have a separate lexical store specialised for reading and that all reading responses are drawn from this store rather than a general lexicon. Further support for this view comes from the observation that children tend to make no response rather than guess at a word which they have not read before (Seymour & Elder, 1986). This suggests that the child knows when a word is not a reading word, for example when it does not look familiar, and refuses to make a response. To be able to make this decision, the representations of reading words must be stored in a separate reading lexicon, or otherwise distinguished from non-reading words in a general lexicon.

The foregoing observations indicate that children's initial representations of print words involve more orhtographic (letter) knowledge and more analytical procedures that appear to be the case in Frith's (1985) Logographic stage. In particular, our hypothesis is that the development of subsequent phonological reading procedures depend on the presence of at least some level of lexical orthographic knowledge. This lexical orthographic knowledge enables the child to later learn about the relationship between letters and sounds within words. These sublexical relations are used by children within the first 12 months of instruction even before any explicit teaching of letter sound correspondences, if this is provided (Thompson & Fletcher-Flinn, 1991). Lexical orthographic representations include some information about the position or ordering of letters within words. Without such representations, it would be impossible for the child to acquire letter-sound associations as the temporal order of sounds must be matched to an ordered, or at least partially ordered, representation of the print word in order to extract useful associations.

The Model

We now present a design for a computational model based on the above constraints and on a consideration of further development towards a mature reading procedure. The design emphasizes the functional properties of the reading procedure with a view to modelling the qualitative performance of children reading at this initial stage. Our methodology in evaluating the model will be to build a number of variations on the central design and to evaluate each variation. This will hopefully give a picture of how the various parameters affect performance and may enable us to chart a developmental course through this first stage of reading. Since our model assumes that children use orthographic, letter based, representation of words the term *Logographic* as used by other writers seems inappropriate as it implies that words are treated as wholes. For this reason we use the term *Visual Reading* to emphasise that word recognition is based on visual cues such as letters and letter clusters rather than being phonologically mediated.

Note that we are modelling only a very early stage in reading development which may last for only a few weeks in some children or a few months in others. We feel that it is important to study this part of development since it tells us about where reading skills come from and how pre-existing knowledge and skills are used by the child to develop a reading procedure.

Lexical Representations for Visual Reading

The model consists of two basic kinds of object: the representation of individual words (both of words in the lexicon and of stimuli to compare with words in the lexicon), and the lexicon itself which stores words and allows them to be accessed in an appropriate way. The form of these objects is related to the algorithms which will be used to manipulate them; these are described in the next section. The representations of both stored words and stimuli have exactly the same form in this model; consequently the description of the representation of words also applies to the representation of stimuli.

The representation of words in the visual lexicon is letter based. Each entry in the lexicon, representing one word, consists of a set of letter tokens corresponding to the letters in the word. This lexical representation may be incomplete as the child may not pay attention to all of the letters in a word in the early stages of learning to read. Consequently a lexical entry may contain fewer letter tokens than the print word it represents. It may also contain letter markers which correspond to letters matching some visually defined criteria; for example, a 'ball and stick' letter (*p,b,d*) or a letter without ascender or descender (note that it is not allowed to contain non-visual markers like *vowel*). Introducing these elements into the representation language allows the model to store an imprecise description of a print word. Such descriptions may help account for children's error responses. This representational framework also allows a word to be represented in terms of non-letter based features, although this is not implemented in the current model.

The set of letter tokens or markers is stored as a partial order. In a partial ordering, the relative order of some pairs of elements in the set are specified while other orderings can be left unspecified. It is therefore possible to record anything from an unordered set of letters to a totally ordered sequence. For example, the partial order [a → b, c, a → d] says that both *b* and *d* come after *a*, but the position of *c* and the relative positions of *b* and *d* are not specified[2]. This partial order could match a number of sequences of letters including: *abcd, acbd, cabd* and *cadb*. A partial order on a set of letter tokens is referred to as a *partial sequence* in the rest of this thesis. As an example of a partial sequence representation, consider the representation that the first author has of *receive* (Figure 1). The relative ordering of *e* and *i* is not specified, but their position between *c* and *v* is. This often results in his incorrect spelling: *recieve*.

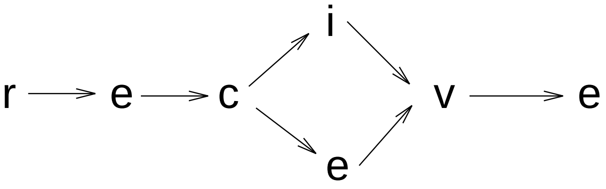

Figure 1. Partial order representation of *receive*.

The visual lexicon is a collection of partial sequence representations of print words organised so that candidate matches to stimuli can be selected quickly. Lexical items are accessed via *keys* which are some feature or features of the word chosen as salient by the reader. Each key designates a subset of the lexicon: a group of lexical items for which the key is salient. This structure enables the word recognition procedure to identify candidate matches by selecting the lexical items which share salient features with the stimulus. Using more than one feature means that the candidate set should be smaller, as only lexical items sharing *both* features with the stimulus will be selected. Examples of salient features might be the first and/or last letters in a word. In this case one key might select all words begining with *b* while another selects those words ending in *t* (see Figure 2) resulting in a candidate set containing

---

[2] A partial order is written here as a list of pairs or singletons separated by commas between square brackets. A pair *a* → *b* means that *a* comes before *b* in the ordering. A singleton element is neither before or after any other element.

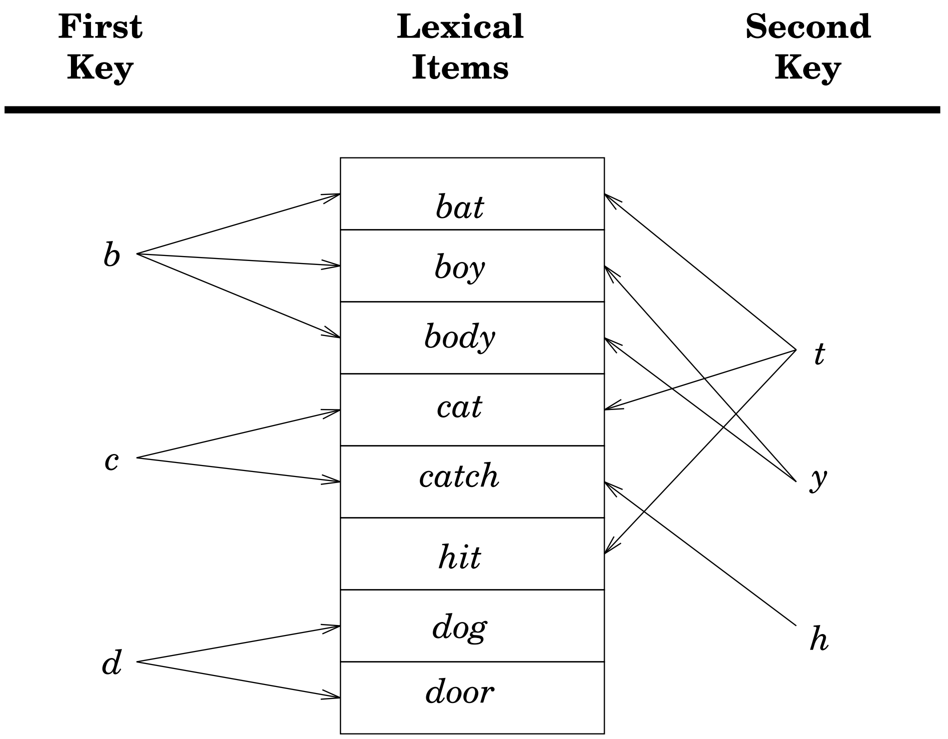

Figure 2. The organisation of the visual lexicon showing the use of first and last letters of stored words as keys. Each key selects a subset of the lexical items.

only words beginning in *b* and ending in *t*. In all of the simulations reported here, the keys have been letters. However, the structure described could be used with keys based on whole-word features such as length (short, medium or long) or outline shape. It is envisaged that such features may well precede the use of letter based features as modelled here.

This lexical structure is similar in concept to Forster's (1976) access file model. The selected keys pinpoint the word's address in the access file, this address is used to retrieve a small set of candidates which, in Forster's model, are then searched serially for a match with the stimulus. Our model goes further in describing the nature of the representation of the individual lexical items and the mechanism by which words are indexed in the 'access file'.

Processes of Visual Reading

Using the structure described above, visual word recognition is achieved by a process which compares a representation of the stimulus with stored representations of lexical items. This section describes that process in detail. In some cases it is necessary to distinguish between the underlying theory being modelled and the implemented model. The differences are due to the requirement that the model run on a computer and provide a usable interface, and to the fact that the model does not implement all of the variations possible within the theory.

The processes of visual reading are depicted in Figure 3 as a 'blackboard' system (Erman, Hayes-Roth, Lesser, & Reddy, 1980) where processes communicate with each other by posting messages at different levels of a central 'blackboard'. The same processes could equally well be shown as a flow chart or data-flow diagram. The processes and their operation are described below.

*Building a partial sequence representation.*Retinal images are first transformed into an internal description that encodes marks on the page as visual perceptual codes (although there may not be a one-to-one correspondence between the elements of this code and letters) in some spatial arrangement.

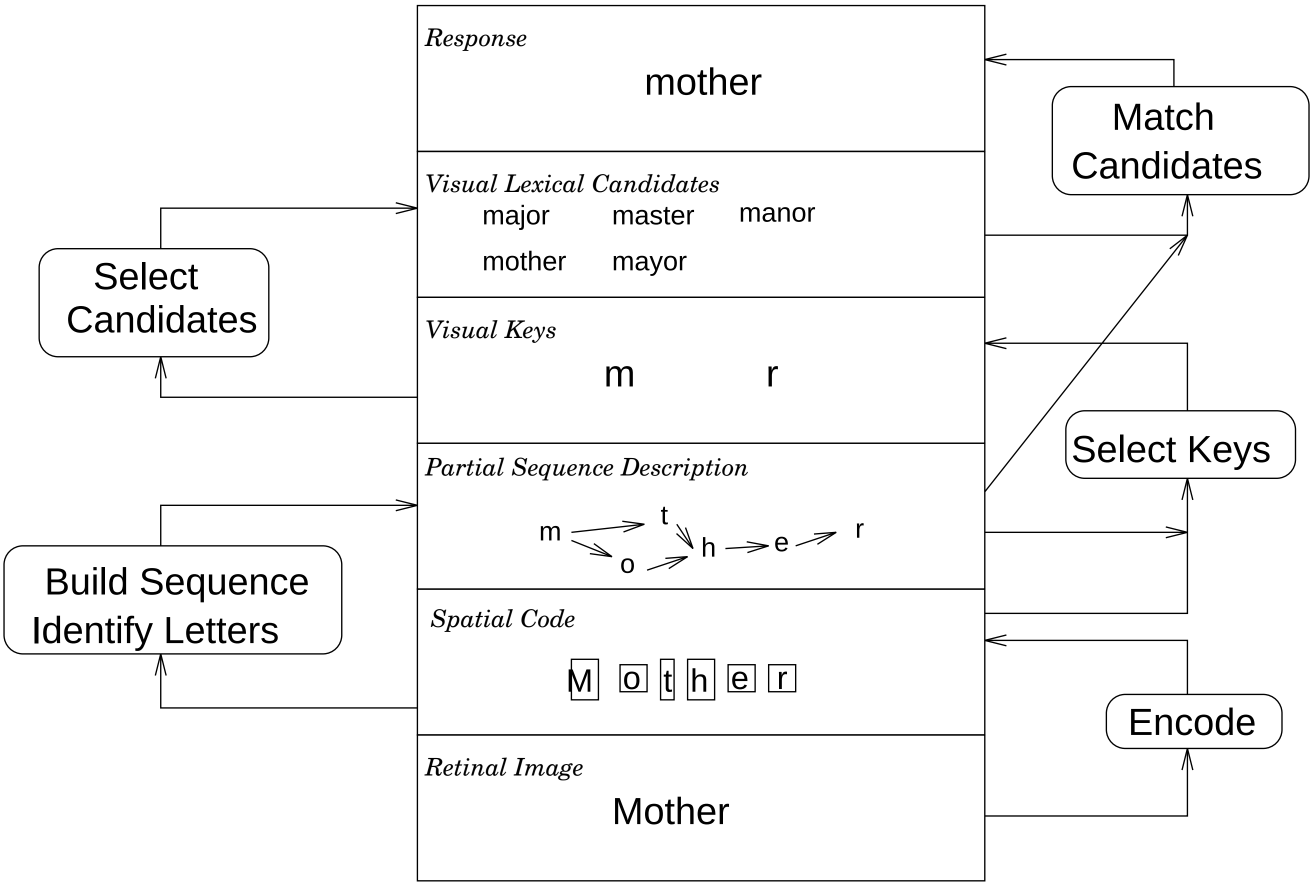

Figure 3. The processes and data flow in visual word recognition are shown here as a blackboard system. Each process reads data from one area of the blackboard and writes its results to a new area.

This *spatial code* may be built up over a number of saccades — not necessarily during one eye fixation — and so is not just a redescription of the pattern of images on the retina. This process is not necessarily specific to word recognition: it may be shared with general object recognition.

This spatial code serves as input to a reading-specific process that extracts some of the information to produce a representation of the stimulus as a partial sequence of letter instances. Letter identities are drawn from a store of graphic letter knowledge. This process builds a more abstract description of the stimulus that serves the purpose of lexical access. Information will be lost in this transformation, such as the exact spatial relationships between the letters in a word. The degree of ordering represented in the partial sequence depends on the stage of development of the child.

This process is modelled here by a procedure which maps a string of letters onto a sequence of tokens and another procedure which then generates a partial sequence description. This procedure is intended only to model the *output* of the initial stages of word recognition, no claim is made about the detailed correspondence between the mechanism used here and that used by children. The first procedure, ENCODE (Figure 3), is expressed as a set of rewrite rules. The procedure may combine letters into one token (for example, the sequence *oo* might be ‘seen’ as one token) or omit letters altogether. The rewrite rules are context sensitive and so can reference the beginning or end of the word or salient visual features in the word such as ascenders or descenders. The details of the procedure used is one of the parameters studied in the simulations reported later in this chapter. As an example of the sort of procedure that can be implemented, the *omit* encode procedure generates letter tokens for the first and last letters in a word and any isolated ascender or descender; all other letters are omitted from the representation. The model's knowledge of letter identities is implemented as a table of input letters and their corresponding internal forms. This allows the model to map some letters to markers (for example, low frequency letters without ascenders or descenders such as *v,w,x,z* might be mapped onto the marker *small*) or to map upper and lower case letters either to the same or different tokens.

A second procedure adds ordering to the letter-based representation. As has been mentioned, the degree of ordering is a parameter for experimentation. The two options investigated here are no ordering, which simply returns a list of the letter-tokens in the word, and boundary ordering. Boundary ordering identifies the boundary and salient letters in the word and orders all letters relative to these. This results in a partial ordering with only one source (start letter) and only one sink (end letter). The relative ordering of the internal letters is likely to be under-specified. Salient internal letters are ascenders or descenders; these letters will have their relative ordering specified and other letters will be ordered relative to them.

*Selecting a key.*The spatial code is also one source of information for choosing the keys which will be used to index into the lexicon. For the early reader, these keys are visually distinct items in the stimulus, such as the first or last letters or isolated ascenders; these items may be the only things represented in the spatial code at this stage. This is implemented in the model by selecting a random element of the spatial code of the word. This fulfils our requirements as only visually salient letters will be recorded in the spatial code in the early stages. Later in development, as the representation of words becomes more accurate, the more abstract partial sequence description may also be used as a source of information in key selection, enabling, for example, the first or last letters to be used as keys. It is envisaged that key selection will become a rule-like procedure operating almost exclusively on the partial sequence representation and using extensive knowledge of orthographic structure. The use of the spatial description for key generation is seen as being inherited from more general object recognition procedures.

The description of the stimulus that results from the two stages of processing described so far is a partial sequence of letter instances and a set of keys. This description is not necessarily complete in that it could omit some letters from the stimulus or may contain instances of markers such as *ascender* rather than letter identities like *h*. For example, the following might be a representation for *little*:

```
key = l
[inst(L1,l) → inst(L2,ascender) → inst(L3,l)]
```

`inst(L1, l)` denotes an instance of the letter `l`. `inst(L2,ascender)` denotes an instance of the class `ascender` (letters with ascenders) called `L2` representing the doubled letters *tt* which have been classed as a single token by the encode procedure. The second instance of `l` is differentiated from the first by virtue of an internal identifier (`L3` rather than `L1`). The partial order contains the information that an instance of `l` is followed by an instance of `ascender` which is followed by an instance of `l`. The letters *i* and *e* are not included in the representation. Once the stimulus is described in this way, it can be matched against candidate members of the visual lexicon, which are also stored as partial sequences.

*Selecting a set of candidate responses.*The key chosen from the stimulus defines a subset of entries in the visual lexicon (the entries referenced by that key). If there are a number of keys, the set of candidate lexical items will be those that are in *every* set indexed by a key. (In fact, the strict intersection of the subsets might not be what is needed since, if an incorrect key is selected, some of the subsets may not contain the target. A better version might be realised as a fuzzy intersection operator; for instance, selecting the lexical items that occur in the majority of the subsets.)

There are a number of algorithms that could realise this indexing mechanism. To make them efficient, indexing (subset selection) and set intersection would have to be primitive or near primitive operations in the functional architecture of the cognitive system — that is, they should be fast, preferably operating in time independent of the number of members of the set. The order in which the subsets are selected is not important and so could be done in parallel. A parallel implementation might also be able to combine the subset and intersection operations effectively to produce a faster procedure.

*Choosing a response from the candidate set.*The set of candidates selected by the indexing procedure may contain zero, one or many items. If there are no candidates, then there is no response for this stimulus; depending on the situation, the reader could either do nothing, or add what may be a new print word to the visual lexicon. If there is just one candidate and if the candidate matches the stimulus representation a response can be made (using the links from the visual representation of the lexical item to semantics and/or phonology). If there are a number of candidates, one must be selected by matching the stimulus against the stored representations. Since the stored representations and the stimulus are both partial sequences, it may be possible to match them a number of different ways. In particular, more than one candidate may match the stimulus. In this event it may be possible to use some measure of 'goodness' of match to select between the candidates. Failing this, an arbitrary choice can be made.

In the implemented model, candidates are retrieved via the key under which they are stored (only one key is used in all of the simulations reported here) in order of recency of access. This means that more frequent words will tend to be examined first when a stimulus is being matched against the lexicon. The model implements a serial search through the candidate set, the first match found identifies the stimulus and determines the response. Due to the imprecise nature of the partial sequence representation it is possible that a stimulus will erroneously match the wrong lexical item; this will result in an error response as the process has no way of knowing that the match is erroneous. The problem of more than one candidate matching the stimulus does not arise in the implemented model as the first matching candidate is always taken as the correct one.

*Adding new lexical items.*When a stimulus is not recognised, the model assumes that it is a new word, not represented in the lexicon. Due to the imprecise nature of the representation, the word may in fact be in the lexicon. For instance, if the model distinguishes between *T* and *t*, then *The* may not be identified with the stored representation of *the*. In addition, many words will be stored under multiple keys if a random key-selection procedure is used as it is in some simulation experiments on this model. This being the case, any strategy used for adding new lexical items will result in a duplication of visual representations of some words. As the form of the stored representation is the same as that of the stimulus, the partial sequence representation of the new word can be added as-is to the lexicon. In order that the new lexical entry be used in recognition it must be indexed under some key. The key chosen for the new item should ensure that it will be amongst the candidate set the next time that this word is a stimulus. The only sensible choice is the key that was just used — the one generated from the stimulus by the procedure described above. Once the new word is installed, further instances of the word can be recognised by comparing them with this new representation.

Alternative Models

In designing this model of visual reading, a number of options were considered and rejected before the final structure described above was settled on. The following paragraphs describe two of these with some discussion of why they were not considered suitable.

*Neural Networks.*Many recent cognitive models, including an influential model of word recognition (Seidenberg & McClelland, 1989), have been built around the connectionist or neural network paradigm. The motivations for the use of these types of models are two-fold. Firstly biological plausibility is claimed as the models work by doing computation in a massively connected network of simple processing units, analogous to the neurons in the brain. Secondly, connectionist models exhibit interesting behaviour in that they are able to learn input/output mappings from examples and generalise to new cases. They seem to escape from the rigid rule-following behaviour of more traditional models (Rumelhart & McClelland, 1986).

The main problems with current connectionist models are representational. The basic language available to a network is propositional: features or entities are either present or absent, true or false, although it is possible to encode degrees of truth in the level of activation of a node. Such a language is found wanting when it comes to expressing relations amongst entities (Pinker & Prince, 1988) such as those that are needed to represent a sequence of letters. Current work in neural networks is finding some solutions to these problems (Elman, 1988, 1989; Jordan, 1986). However, they remain problems with the current state of the methodology.

One of the goals of this project was to develop a functional specification for the representations and procedures used by a child when learning to read. It was felt that the constraints of current connectionist models were too limiting in this respect. Models built using connectionist techniques are not free to choose an appropriate representational structure; the form of representation is dictated by the implementation technology rather than the cognitive task under consideration. A symbolic model on the other hand can be a more direct implementation of a functional description and thus allows us to verify that a functional description is valid. Once the model is specified in this way, a connectionist network may turn out to be a more appropriate implementation of all or part of the model.

*Spreading Activation.*Rumelhart and McClelland's (1982) model of the word superiority effect used a mechanism known as spreading activation. The model consisted of a network of nodes representing letter features, letters and words. The letter features observed in the stimulus were 'activated' and this activity then spread to the letters which contained those features and the words containing the letters. The activity of a letter node depended on the activity of its letter features and similarly for word nodes. Word nodes inhibit each other with the result that only one word node will tend to be active after some settling time. This mechanism has also been used in a number of cognitive models and forms part of the ACT* cognitive architecture proposed as a general model of cognition by Anderson (1983).

Spreading activation is a way of finding the most appropriate match for a given set of observed features. Competing hypotheses at any level will inhibit each other with the result that the hypothesis with the most supporting evidence, in the form of activation flowing from other nodes, will win out. A major problem with spreading activation is again representational. Although the network through which activation flows is allowed to have arbitrary relations, such as *part-of* or *next-to*, their semantics with regard to activation flow is fixed: they either excite or inhibit. Nodes are again propositional and so it is hard, for instance, to represent a word with two *t*'s without having 26 nodes for each possible letter position.

The difficulties with propositional networks was a major barrier to the use of spreading activation in this project. However, considered as an abstract process, spreading activation can be seen as a special sort of indexing scheme where a candidate set is selected by being activated by a set of keys. The first key, say initial letter *m*, activates all words beginning with *m*, the second activates a subset of these until there is only one or a small number of candidate words. Consideration of this sort of selection algorithm led to the model of the visual lexicon described above.

Simulation Experiments and Evaluation

The main reason for implementing a model is to explore the consequences of the theory on which it is based. The model of visual word recognition described here can be used to find out whether the suggested mechanisms can give rise to the behaviour observed in young children when they first begin to read. Without the model, it is hard to tell whether the theory 'works' — whether it is capable of recognising words or distinguishing a sufficient number of words. This section presents a number of simulation experiments which show that the model is capable of appropriate behaviour given a reading environment which reflects that of a young reader.

Simulation Method

The basic method for evaluating the theory/model is to present the model with a corpus of words to read and learn. In its normal 'training' mode, the model is set up so that when given a word it will attempt to read it (match it against a stored representation) and if it fails, will store it in its lexicon. For evaluation, the model can also be run in 'reading' mode when it will not add any new words to the lexicon. In the simulation experiments, a number of parametric variations on the model are trained on the words from a small set of stories. The performance of the different variations are compared to each other and to observations of children's performance at the start of their reading development.

The training vocabulary used in these simulation experiments is drawn from books used in the first year of reading instruction in New Zealand schools. All of the books used in the first two years of teaching at one Wellington school have been transcribed into a machine readable format. The books used for these experiments are part of the Emergent level of the *Ready to Read* and associated series which would be used in the very first phase of reading instruction. The training set consists of twenty stories which contain 677 tokens (individual word instances) and 233 types (different words). The model is given these words in the order that they appear in the stories, so word frequencies are as they would be for the child. During training, each story is presented to the model twice, one after the other. On the first presentation, many of the words will be new and so cannot be recognised but will be stored in the lexicon. On the second presentation, all of the words will be 'familiar' (since the model's memory is perfect); any errors then reflect real recognition errors rather than an unfamiliarity. Most of the evaluation of error performance is done on this second presentation. To aid in the evaluation of the model, an additional word list was constructed consisting of 45 words from other Emergent level books from the same series books not contained in the training set. This was presented to the model at the end of training to assess its performance on unfamiliar items.

When the model is given a word, it either responds with a word from its lexicon or indicates that no lexical item was found to match. When a word response is made this could either be the same as the input word, in which case the response is correct, or not, in which case a substitution error has been made. Thus there are three response categories: correct, substitution error and refusal. When a substitution error response is made, the target and response are saved for later analysis.

Parametric Variations on the Algorithm These simulation experiments are an exploration of the capabilities and assumptions of the implemented model. There are a number of parts of the model which admit to some parametric variation: different versions of some procedures can be realised within the specification of the theory. To evaluate the theory/model, a number of variations were constructed and trained in the manner described above.

The parameters of the model reflect both different assumptions about the form of internal representations and different levels of knowledge of the reader. In some cases the different representational assumptions can be seen as points on a developmental progression: for example, the increase in precision of the output from the ENCODE procedure. The following list details the parameters varied in these simulation experiments.

Spatial Encoding: The ENCODE procedure produces a tokenised description of the print word. It is possible for this procedure to miss out some letters; the variations concern what happens when letters are missed:

- OMIT: This procedure is intended to simulate the child who only attends to and retains a few letters in any word. Only boundary letters and some salient internal letters (isolated ascenders or descenders, or doubled letters) are retained. All other letters are omitted. The rationale for this is that these are the visually distinctive elements of the letter array and so will be retained as part of a visual representation of the stimulus.

- MARKERS: This procedure simulates a child who sees most of the letters in a word but only identifies some of them. Most (not all) non-salient internal letters are replaced by markers showing the class of each letter (eg. ascender) but there is a small probability that internal letters will be identified properly. This procedure is intended to retain some implicit length information.

Letter Knowledge: The accuracy with which the ENCODE procedure is able to identify the letters from marks on the page. Two levels of accuracy are included:

- NOVICE: Upper and lower case letters are differentiated, *b* and *d* may be confused, *f,g,j,q,x,* and *z* are identified as *ascender*, *descender* or *small.*
- EXPERT: Upper and lower case letters represented as the same token, all letters except *x* and *z* identified reliably.

Ordering: The BUILD PARTIAL SEQUENCE procedure identifies the tokens as letters and records their ordering. The variations concern the degree of ordering recorded:

- NONE: No order is recorded
- BOUNDARY: Everything is ordered relative to the boundary letters although the relative ordering of internal letters/markers is not recorded

Key: The GENERATE KEY: procedure chooses some feature of the word token to use as a key for selecting the appropriate subset of the lexicon. In the current experiments there are only two versions:

- RANDOM: A random letter is selected from those included in the encoded version of the word token. This procedure won't choose markers, such as *ascender*, as keys.
- INITIAL: the first letter of the word token is used.

Six different combinations of parameters were used in simulation experiments. For discussion, simulations are named by an abbreviation of the three main procedures that were used, for example *ma.bo.ra* designates the *MArker* encode procedure with *BOundary* ordering and *RAndom* key selection. Since the non-ordering simulations were assumed to model an earlier stage of learning, the *Novice* level of letter knowledge was used for these, all others used *Expert* letter knowledge. In later simulation experiments the effect of the level of letter knowledge on performance is investigated further.

Collection of Results When the model is being trained, records are kept of the responses to each word in the training set. These records can then be used to analyse a number of aspects of the model's performance. Unless otherwise stated, all measurements were made during the second presentation of each story. The simplest measure is a count of the number of each kind of response: correct, substitution error and refusal. These were counted on a per-story basis as well as on the training set as a whole. In addition to monitoring the responses during training, which for the most part reflect performance on familiar words, the model was also given a list of unfamiliar items to read at the end of training. The model could not possibly get any of these words correct as they are not in its lexicon. The results show which sort of error, substitution or refusal, is more likely for an unfamiliar word.

As well as collecting error data, some statistics on the internal state of the model were also collected as training proceeded. The size of the lexicon at the end of each story was recorded. As was mentioned in the description of the model, multiple representation of words is quite likely when the representation procedures are unstable (don't give the same results every time). The size of the lexicon relative to the number of distinct words actually stored is a measure of the level of redundancy in the lexical representation.

Another indicator of the state of the lexicon is the number of candidates selected when matching against a stimulus. A large mean number of candidates would indicate that lexical items were inefficiently distributed amongst keys (eg. many items stored under 'e' and few under 'f'). In fact, in all of the simulations, this measure was found not to provide any interesting information: mean number of candidates was simply related to lexicon size in all cases.

Simulation Results

Table 1 presents the error rates in the six simulations. The table gives the proportions for each response type, averaged over the whole training run, and the percentage of refusal responses to the unfamiliar word list.

Experiment	Familiar Words Correct	Familiar Words Refusal	Familiar Words Substitution Error	Unfamiliar Words Refusal
om.no.ra	76	9	15	63
ma.no.ra	59	29	12	82
om.bo.ra	74	17	9	65
ma.bo.ra	61	36	3	78
om.bo.in	85	5	10	63
ma.bo.in	75	21	4	80

Table 1
Results of simulations of visual reading, percentages.

Quantitatively, the performance of most of the simulations is reasonable; the model gets most (59 to 85%) words right and refuses on 5 to 36%. As a point of comparison, Table 2 presents some results for four beginning readers reading familiar and unfamiliar words taken from Seymour and Elder (1986). The performance of different children is quite varied but shows a pattern of more refusals than substitution errors, especially in the good readers. The children tend to make no response to an unfamiliar word rather than making an incorrect response. This suggests that the representation of the input stimulus is such that it will only match a very similar lexical item; that is, that the representation and the matching procedure are reasonably precise. The model, in most cases, also tends to make refusals rather than substitution errors to unfamiliar words (Table 1). This comparison is not meant to show that the model's performance mimics that of any one child. Seymour and Elder presented results from a large number of children, the ones selected here are just a small sample. Children are subject to individual differences not only in reading skills but in general language ability and general intelligence, all of which will affect their performance. In addition, some of Seymour and Elder's subjects seem to be using another reading strategy as they are able to name at least some 'unfamiliar' words. All of this means that we can only make a 'ball-park' comparison of our model's performance with that of children, looking at the pattern of the results rather than the exact numbers.

	Familiar Words			Unfamiliar Words		
	Correct	Substitution Error	Refusal	Correct	Substitution Error	Refusal
KN	79	4	18	7	6	88
CR	72	8	20	1	8	90
NW	63	20	17	3	24	74
DR	38	14	48	2	18	81

Table 2
Percentage error performance of four beginning readers (taken from Seymour and Elder (1986).

Table 3 shows the correlation (Pearson's product-moment correlation) between the length (in letters) of the target and the length of the response for all substitution errors made in the simulation experiments. Seymour and Elder (1986) suggests that the size of this correlation indicates that children use length as a cue in early reading; they observes correlations as high as 0.9 in first year readers reading words in isolation. The partial sequence representation used here does not contain an explicit representation of the length; rather it is implicitly encoded in the number of letter tokens in the word. It can be seen that the degree of correlation between target and error is a function of the amount of information lost by the spatial encoding procedure. Thus the *omit* procedure gives a non-significant negative correlation and *marker* give a correlation of up to 0.51. It is not hard to see why this should be the case: the *marker* encoding procedure represents length (using markers) more precisely than it does the identity of the letters that make up the word; any error must be similar in at least its length to the target, otherwise they will not match. The fact that Seymour finds such a large target/error correlation might be accounted for by a marker encode procedure which is more precise and produces a marker for *every* letter in the word, rather than missing some letters as it does in these experiments. Note that the encode procedure cannot be much more precise in recording letter identities as this would mean that almost no substitution errors would be made. Clearly a balance point must be found to be able to model both results.

Experiment	Target-Error Length Correlation	Lexicon Size
om.no.ra	-0.13	1.49
ma.no.ra	0.45	2.30
om.bo.ra	-0.03	1.82
ma.bo.ra	0.35	2.65
om.bo.in	0.00	1.25
ma.bo.in	0.51	1.94

Table 3
The correlation between target and error lengths for substitution errors in six simulation experiments. The Lexicon Size column shows the final size of the lexicon expressed as a multiple of the optimal size.

As a measure of the efficiency of lexical storage, Table 3 shows final number of lexical items divided by the number of types (distinct words) represented in the lexicon. For efficient storage, this ratio ought to be one, so that any word has just one representation. As has been mentioned before, the model may store a word a number of times if the key selection procedure or the stimulus encoding procedures are not consistent — that is, if they don't give the same result each time. The table shows that all of the experiments produced more than one representation for many lexical items. The marker encoding procedure is particularly variable in the letters it leaves out and so tends to produce a number of similar

om.no.ra	ma.no.ra	om.bo.ra	ma.bo.ra	ma.bo.in
Where/We	hook/hill	Baby/bumpy	Four/for	He/Here
after/zebra	jet/went	He/house	Here/house	Not/net
am/zebra	keys/silly	Not/net	Not/net	Six/swim
bench/hill	lion/climb	Sam/swim	Sam/swim	These/table
bike/broke	lunch/hole	Sam's/Six	Six/sun	These/the
builder/roll	motor/storm	These/The	These/the	These/time
by/birthday	nurse/comes	Three/tree	Three/tree	What/went
cook/cake	roller/motor	Who's/What's	Two/to	Who's/wears
fast/boat	sat/this	bam/Bim	bam/Bim	bed/bird
floor/door	school/hole	bed/bird	bed/bird	broke/bike
hair/roll	sun/comes	by/bumpy	fire/Five	fire/Five
home/house	swim/comes	fire/Five	for/Four	home/here
nose/nurse	tall/went	just/jet	home/Here	time/table
right/tiger	then/net	made/Me	table/the	time/the
sat/this	this/out	nose/nurse	the/time	too/took
table/the	time/went	run/rain	time/the	wears/What's
tall/boat	too/the	table/The		will/wall
they/jet	very/cowboy			
to/out	wearing/cowboy			
too/the	wide/ride			

Table 4
Substitution errors made in some of the simulations, Target/Response.

but incompatible representations of the same word.

Some examples of the substitution errors produced by the simulations are shown in Table 4. Note that many of the errors for the unordered simulations (*om.no.ra* and *ma.no.ra*) share only one letter with the target. Also, there are a number of cases where the position of the common letter is 'reversed' — it appears at the start of the target and the end of the response. When ordering is introduced this feature disappears altogether. The rest of the table is offered without comment as examples of the behaviour of the model.

The effect of introducing ordering to the representation is not clear from the results presented above. With both the *omit* and *marker* encode procedures ordering reduces the number of substitutions, and for the *marker* procedure, increases the number of correct responses slightly. One problem with the experimental design here is that ordering is confounded with level of letter knowledge since the non-ordered simulations used Novice letter knowledge whereas the ordered simulations used Expert letter knowledge. The effect of this on performance will be discussed later. The main difference between ordered and non-ordered cases is in the nature of the errors produced and their relationship to the target. Table 5 shows the proportion of errors sharing first or last letters with the target in each simulation. When ordering is introduced the proportion of errors sharing first or first and last letters with the target increases while the proportion sharing last or some internal letter declines. This results from two factors: the preference of all of the spatial encoding procedures for boundary letters and the introduction of ordering which precludes a match between items sharing letters in different positions. This data clearly relates to that presented by Stuart and Coltheart (1988) which showed a similar pattern of errors in begining readers. This data will be discussed in a later section.

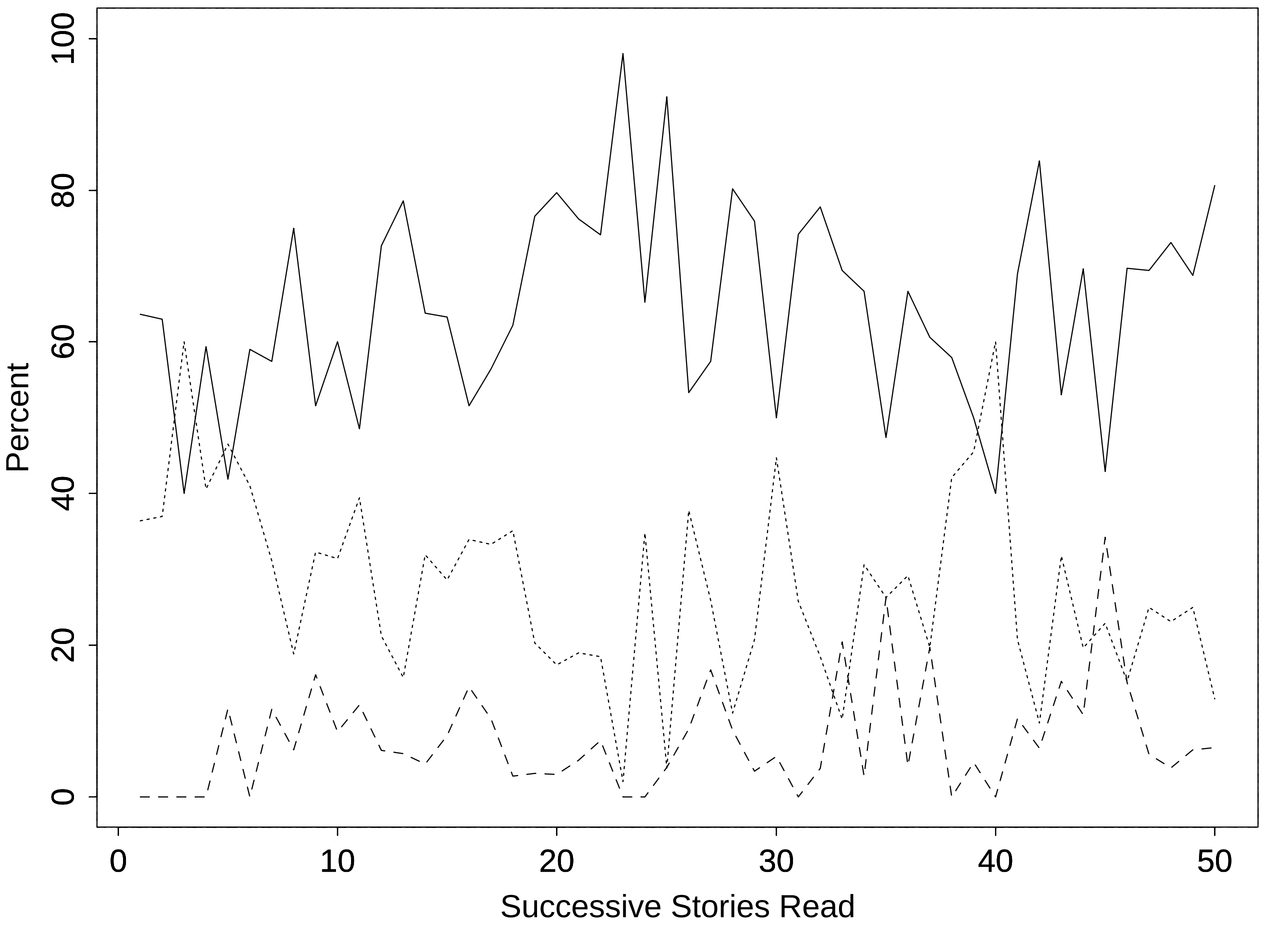

Figure 4. Proportion of correct (top, solid line), null responses (middle, dotted line) and substitution (bottom, dashed line) over 50 stories for the *ma.bo.ra* simulation. Scores are in percentages measured for each story read by the model.

Simulation	N	First & Last %	First %	Last %
om.no.ra	231	39	21	14
ma.no.ra	197	4	25	19
om.bo.ra	143	89	11	0
ma.bo.ra	48	98	2	0
om.bo.in	156	98	2	0
ma.bo.in	71	80	20	0

Table 5

Letters shared between substitution errors responses and the target

The Capacity of the Visual Lexicon The visual lexicon, as described above, is capable of storing, and differentiating, a large number of words. However its performance changes as the number of words increases. To examine these changes, one of the experimental configurations (*ma.bo.ra*) was run on 50 stories, containing 493 types and 1985 tokens (taken from the Emergent and subsequent Early levels of the book series). Figure 4 shows the change in the proportion of correct and error responses and refusals over the course of reading 50 stories. A peak of performance is reached after around 25 stories with over 80% correct responses and a small number of substitutions. On reading more stories, the number of correct responses drops to around 60% although after 50 stories it does seem to be rising again. As can be seen from the graph, there is a large amount of variation in scores between successive stories. Most of this is due to the variation in vocabulary between stories: some words will tend to cause more errors than others.

The major change in the lexicon as more words are added is that more time must be spent in the linear search within a set of candidates. Whereas after 20 stories there are on average 39 candidates for each key, after 50 stories there are 80. This means that the recognition procedure spends most of its time in a linear search of the candidate set, matching each candidate against the stimulus representation. The larger number of candidates increases the probability that an erroneous match will be found, resulting in a substitution error; this manifests as an increase in the proportion of substitution errors towards the end of the simulation. The main effect of the large candidate sets is that naming times would increase on average as the size of the lexicon increased — assuming that the search is sequential rather than parallel.

Most of these consequences of increased lexicon size are undesirable, in that they are not exhibited by children learning to read. It is clear that some qualitative change must take place in the way that the visual lexicon is organised. The situation can be improved in two ways. Firstly, because of the way in which the lexicon is built, it contains a large number of duplicate entries. If this redundancy could be removed the number of candidates might reduce by as much as a half in some cases. The second method of reducing the size of the candidate set is to use more than one key to index into the lexicon. If, say, both first and last letters were used, the number of candidates would be a great deal smaller.

Representational Change in Visual Reading

The model presented here learns only in the sense that new words can be added to the lexicon as they are encountered. Evaluation of the simulation results above showed that the basic model was consistent with a number of results from research on child reading but that some combinations of parameters were more successful than others. Based on the findings presented above, and on further evaluation to be discussed here, the following developmental sequence is proposed for the visual lexicon:

Phase 1: The initial word recognition procedure uses the *marker* encoding procedure, records no ordering in the letter array and chooses a random visually salient element of the letter array as a key. Allographs of letters may be differentiated, eg. upper/lower case, some letter confusions occur. (Simulation *ma.no.ra.*)

Phase 2: Ordering information is introduced into the representation of the letter array with internal letters ordered relative to the word boundaries. Almost all of the letters can be identified consistently independent of case. (Simulation *ma.bo.ra.*)

Phase 3: First letter (or perhaps first and last letters) is used as a key for the retrieval of candidate lexical items. (Simulation *ma.bo.in.*)

This sequence provides for an increase in accuracy of word naming from 59% in phase 1 to 75% in phase 3 while maintaining a significant correlation between target and error lengths. In all cases

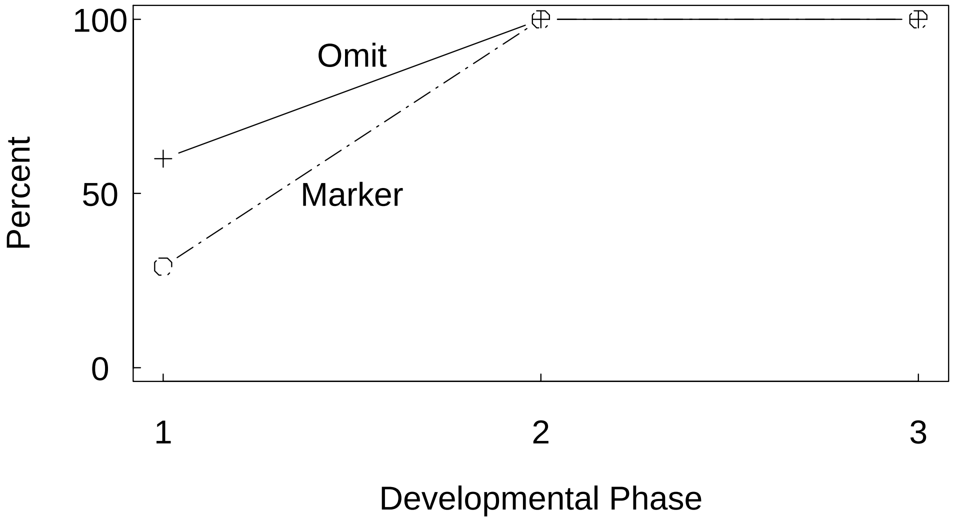

Figure 5. Percentage of all substitution errors sharing first or first and last letters with the target for two sequences of three simulations

unfamiliar words are rejected about 80% of the time rather than resulting in substitution errors. Each phase is qualitatively different in the representations and procedures used for word recognition.

We can compare the performance of this proposed developmental sequence with the error data collected by Stuart and Coltheart (1988). As discussed earlier, they found that the type of errors that children made changed as their reading improved. In particular the proportion of errors sharing the first or first and last letters with the target began to predominate when the children passed a criterial test of phonological awareness. To evaluate the model's performance in this respect we can classify the substitution errors made in each of the three marker simulations which make up the developmental sequence discussed above. This data, presented in Table 5 is summarised in graphical form in Figure 5.

It can be seen from Figure 5 that the proportion of this type of error is always above 50% when the *omit* encode procedure is used. This is because in this case the first and last letters are perhaps the only letters represented in a word and therefore will be shared between error and target. In the marker case, when no ordering is assumed (phase 1) the proportion of these errors is small but rises to 100% when ordering is introduced. An intermediate point on the marker encode curve can be generated by running the *ma.bo.ra* simulation again with the *Novice* letter knowledge used in the *ma.no.ra* simulation. The result of this change is that some letters are not identified reliably and so confusions can occur, for instance between *b* and *d*. This new simulation yields 66% substitution errors sharing first or first and last letters with the target (Table 6). Clearly the child's letter knowledge affects the proportion of errors in this category: the more accurate the child becomes at recognising letter shapes, the more errors will tend to share first or first and last letters with the target. However, better letter knowledge cannot account for all of the increase in these errors, independent of the use of ordering: when the *ma.no.ra* simulation is run with *Expert* letter knowledge only 38% of errors share first or first and last letters with the target. If we assume then that the child first gains improved letter knowledge and then adds ordering to the representation, the model predicts that the proportion of these errors increases from

Ordering	Letter Knowledge	Number of Substitutions	First & Last %	First %
None(*)	Novice	197	4	25
None	Expert	148	27	11
Boundary	Novice	68	41	25
Boundary(*)	Expert	48	98	2

Table 6
*Letters shared between substitution errors and target for simulations using the marker encode procedure, including two new simulations. The entries marked (*) correspond to experiments reported previously (*ma.no.ra *and* ma.bo.ra*).*

29% to 38% to 100%. If ordering is added before better letter knowledge the progression is 28% to 66% to 100%. In reality, improving letter knowledge is unlikely to be as discrete as modelled here and something in between these two patterns might be expected. The general pattern of these data provides a good fit with the data presented by Stuart and Coltheart, however we still have to account for why these errors begin to predominate when the children pass the phonological awareness tests.

Stuart and Coltheart's (1988) phonological readiness criterion, after which children started to produce more 'phonological' errors, is that the child:

- is able to provide the common corresponding sound for more than half of the letters, and
- shows some success in the following tests of phonological awareness:
  - *Rhyme production:* Subjects produced rhymes for words presented as pictures
  - *Rhyme detection:* Subjects had to differentiate between rhyming and non-rhyming pairs of pictures.
  - *Supply final syllable and supply final phoneme:* Subjects completed the name of a series of small toys, supplying the final phoneme or final syllable where appropriate. For example, experimenter says *Ze*, child says *bra*.
  - *Identify initial phoneme:* Subjects had to identify that the common feature among sets of three objects was their first sound.
  - *Segment initial phoneme:* Subjects had to find the odd one out among a set of pictures based on the initial phoneme.

To achieve any success in these tests the child must have access to a segmented phonological lexicon or a procedure which can achieve segmentation of phonological forms. The segmentation must be able to separate out at least the first and last syllables and phonemes from the word. In one case (supply final syllable/phoneme) access to these segments must be conscious and available to verbal report although in the other cases unconscious access would be sufficient. The hardest tests for Stuart and Coltheart's subjects were those involving phoneme segmentation (supply final phoneme, for example, had a mean score of 2.1 out of 8, relative to 6.6 out of 8 for supply final syllable).

To account for why success in these tests occurs at the same time as the change in substitution error types we need to provide a link between the internal changes in the model and the external behaviour observed in Stuart and Coltheart's tests. The most obvious link is between the letter sound test and the model's visual letter knowledge. These criteria are different in that knowing how to recognise a letter visually does not mean that the child knows its sound. Thus the child is likely to pass a visual letter discrimination test before a letter sound test. However, the *Expert* letter knowledge used in the model corresponds to the child being able to recognise *almost all* letters consistently and so is more strict that Stuart and Coltheart's test which requires that the child knows only half of the letter sounds. It is therefore not unreasonable to assume that a child might fulfil both of these criteria at about the same time.

There is a more complex relationship between the other internal changes in the model and the onset of phonological awareness as observed in Stuart and Coltheart's behavioural tests. To argue that there is a link here we must provide a means by which changes in the child's reading procedure are related to changes in the child's phonological knowledge. This might be seen as problematic here since our reading model is entirely visual; however, we would like to argue that there is a process of representational change at work in the cognitive system which serves to transfer knowledge and knowledge structures from one task domain to another. Such a model of development, dubbed representational redescription, has been proposed by Karmiloff-Smith.

Karmiloff-Smith presents a fairly detailed picture of the role of meta-knowledge in the development of language production (Karmiloff-Smith, 1979, 1986, 1988). She also discusses an example from children's acquisition of drawing skills (Karmiloff-Smith, 1990). Her model revolves around the idea that the first representations that children learn in a given task area are implicit in the sense that they cannot be accessed or manipulated for any other task than the one they were learned for. Once these representations are stable, they are redescribed into a form with exactly the same content, but which is accessible to the rest of the cognitive system. At this time, meta-procedures can inspect the representations and might rebuild them to make them more efficient, generalise them to remove redundancies or any number of other operations which can only be done with a global view of the stored knowledge. An important implication of the model is that it takes developmental time for learned knowledge to become available to other parts of the cognitive system.

Using this model, Karmiloff-Smith presents an account of a frequently observed pattern of learning. The child begins by acquiring a procedure for the task but it is stored in implicit form and works only in one specific case. At this stage a number of similar procedures may be acquired for similar tasks, but since they are all separate, they don't interfere with each other and the child can perform well. Once good performance is achieved, representational redescription begins and, for the first time, the child may see inconsistencies or duplications in the learned knowledge. In trying to resolve these inconsistencies the child's performance on the task degrades. At some later time the representations will have been rebuilt in a more efficient form which allow the same and perhaps better performance than was possible in the first phase. In Karmiloff-Smith's model, the child is driven not only by a goal of mastery over the task but by a goal of mastery over their knowledge and representations used in the task.

For our purposes, we are interested in the relationship between the child's growing phonological skills and the development of the visual reading lexicon. Stuart and Coltheart's phonological awareness tasks all involve manipulation of phonology in a way that is 'unnatural' in the sense that the manipulations don't occur in normal language use. This suggests that the child must have gone through the redescription phase for this knowledge to make it available for manipulation by the cognitive system. So, when the child is able to pass the phonological awareness tests, the phonological knowledge will also be available for integration in some form into other cognitive task areas.

The visual lexicon has become reasonably efficient at identifying words using unordered representations and has perhaps reached a stable point in development. This is the point at which Karmiloff-Smith suggests redescription will take place. Upon examination of the lexical representations, the cognitive system might be struck by the similarity of the representation of some words, *was* and *saw* for instance. Since the phonological lexicon is now 'open' as it were, it might become obvious that one way of differentiating these words would be to include ordering, which will surely be present in the phonological representation. Now some attempt at adding ordering might be possible using embryonic sound to letter knowledge in combination with the phonological lexicon and the existing unordered visual lexicon. Alternatively, the child could just modify the appropriate representation building procedures and rebuild the lexical representations of words as they are encountered.

The change from un-ordered to ordered representations, then, is achieved through representational redescription. It requires that the phonological lexicon has been redescribed into a form that allows external access. Thus the child will not begin to use an ordered visual lexicon until phonological representations are open to manipulation by the rest of the cognitive system. This accounts for the observed correlation between the onset of phonological awareness and the increase in the incidence of errors sharing first or first and last letters with the target. These errors are due to the use of ordering in the visual lexicon.

Even within this framework, there are alternative explanations of the observed pattern. It is possible that the visual lexicon is the source of the use of ordering, rather than the phonological lexicon. In this scenario, the child would first develop ordered representations in the visual lexicon. This might occur as the child redescribes the visual lexicon, noticing that ordering information present in the spatial description is lost in the more abstract letter based description. This structure may then transfer at a later stage to the phonological lexicon when it is redescribed. In this case, the change to the phonological lexicon is above and beyond what is currently needed for spoken word recognition or production. This is consistent with Karmiloff-Smith's motivation for representational redescription: that the child goes beyond behavioural mastery towards mastery over the representations themselves. Once the phonological lexicon has been restructured, the child will be able to pass Stuart and Coltheart's phonological awareness tests. Note that in this case, the behavioural change in reading will precede the change in phonological awareness; the two events are still related and so this account of Stuart and Coltheart's observations may still be valid.

A third possibility exists: that a source external to both the visual and phonological lexicons acts as the impetus for the inclusion of ordering. In this case, the behavioural effects of the change of structure of each lexicon would be expected at different times, depending on when each was redescribed. This might explain the differing results obtained by various researchers attempting to determine the causal relationship between phonemic awareness and reading skill (Bradley & Bryant, 1985; Morais et al., 1979). It is clear that reading skill and phonemic awareness are correlated. If both rely on the use of ordering in the internal representation of words then the order in which the criterial tests are passed may depend on the readiness of each process for redescription (in Karmiloff-Smith's sense). The difficulty in this interpretation is finding a candidate cognitive procedure which could act as the source of the use of ordering in both these lexicons. It must be something which becomes available for redescription at around the right time, since if it is ready long before either lexicon is ready then there is no reason for both lexicons to be redescribed at the same time. This possibility does not seem very credible but is presented here for completeness.

Summary

We have presented a computational model of word recognition in the very first processes used by children learning to read. An important consideration in a model that purports to address this early stage in reading is that its prerequisites should be clearly available to the non-reading child. This is the case in the model presented here which builds on procedures similar to general object recognition procedures. The main requirement is that the child be able to distinguish a number of letter forms from one another, although not all. The model allows for the child's letter knowledge to be extended during this first stage. For average progress readers, this stage would be expected to occur within the first three to four months of school instruction. It differs from the Logographic stage proposed by Frith (1985) in that procedures are more analytic and there is a greater reliance on letters and their order within words. The model in effect describes the acquisition of an elementary visual orthographic lexicon by the beginning reader.

The model uses a lexicon of stored representations of print words as part of the word recognition process. Each print word is represented as a partial order on a set of letter or letter class tokens. The partial order allows the child to represent anything from an unordered set of letters to a complete sequence within the same framework. Individual lexical entries are stored under one or more keys, for instance the first or last letter, such that all of the items stored under one key can be retrieved quickly. Word recognition is achieved by first building an internal representation of the target stimulus. A set of candidate lexical items is then selected using a key derived from the target and each candidate's representation is matched with that of the target. The first candidate to match in this linear search is given as the response. If no match is found, no response is made and the target may be added to the lexicon as a new print word.

Evaluation of the model was based on simulations which varied the parameters of the model. The parameters were concerned with the type of representation of a stimulus that is built and the keys used to organise the lexical store. The results of these simulation experiments was used to define the behavioural properties of the model and a number of observations from the empirical literature were used as points of comparison.

A three phase developmental sequence was proposed for the first stage of reading. In phase one, the child treats words as an unordered set of letter tokens, most but not all of the letters are identified and upper and lower case letters are distinguished. In particular, the boundary letters are more likely to be identified than internal letters: those letters not identified properly are represented by a marker such as *ascender*. A random letter is chosen as the key for candidate selection. Phase two sees the child introducing ordering into the representation such that the boundary letters are uniquely identified; letter knowledge is improved so that most letters can be identified. Boundary letters are still more likely to be identified properly than internal letters and, again, a random letter is chosen as the key. Phase three sees an improvement in the organisation of the lexicon due to the use of the initial letter as the candidate selection key. This reduces the redundancy in the lexicon as words will now only be stored once, and means that candidate sets will be smaller. This three phase sequence provides a credible developmental progression from pre-reading through the first stage of word recognition.

The simulation results were also compared to the error data presented in Stuart and Coltheart (1988) as part of their longitudinal study of a group of beginning readers. They found that substitution errors sharing first or both first and last letters with the target began to predominate at a certain point in the child's development. This point corresponded to the time at which the child was judged to be ready to apply phonological knowledge to the reading task. In the model, using the developmental sequence proposed, the same change in error types is observed. Errors sharing first or first and last letters predominate after ordering is introduced into the lexical representation: phase two and beyond.

An account of the change in error types observed by Stuart and Coltheart was presented based on the ideas put forward by Karmiloff-Smith. The link between the child's readiness to use phonological knowledge and the onset of the use of ordering is explained by the sharing of knowledge structures between the phonological and visual lexicons. This sharing can only take place, according to Karmiloff-Smith, when the phonological lexicon has been redescribed into an explicit form. And this redescription will only occur when the child has achieved mastery in the use of that lexicon. Consequently, the transfer of the concept of ordered representations from phonological to visual lexicons must wait until the child has become phonologically aware.

The application of Karmiloff-Smith's ideas to models of reading development enables a detailed account of the relations between stages to be put forward. In this case the account is based on a computational model built to test some detailed hypotheses about the procedures used in the first stage of reading. The model proved to be a useful medium in which to explore the different possibilities within the current theories and provided the discipline to force every detail of the theory to be stated explicitly. Our attempt to model the first stage of reading development will surely prove to be incomplete but we hope it will provide a starting point for further work in this area.

References

Aitchison, J. (1986). *Words in the mind: an introduction to the mental lexicon.* Lawrence Earlbaum and Associates, Hove, UK.

Anderson, J. R. (1983). *The Architecture of Cognition.* Harvard University Press, Cambridge, MA.

Bradley, L. & Bryant, P. (1985). *Rhyme and Reason in Reading and Spelling.* University of Michigan Press, Ann Arbor, MI.

Campbell, R. (1987). Oral reading errors of two young beginning readers. *Journal of Research in Reading*, *10*(2), 144–155.

Elman, J. (1988). Finding structure in time. CRL Technical Report 8801, Center for Research in Language, University of California, San Diego.

Elman, J. (1989). Structured representations in connectionist models. CRL Technical Report 8901, University of California, Center for Research in Language, San Diego.

Erman, L. D., Hayes-Roth, F., Lesser, V. R., & Reddy, D. R. (1980). The Hearsay-II speech-understanding system: integrating knowledge to resolve uncertainty. *ACM Computing Surveys*, *12*(2), 213–253.

Forster, K. (1976). Accessing the mental lexicon. In Wales, R. & Walker, E.(Eds.) , *New Approaches to language mechanisms.* North Holland, Amsterdam.

Frith, U. (1985). Beneath the surface of developmental dyslexia. In Patterson, K., Marshall, J. C., & Coltheart, M.(Eds.) , *Surface Dyslexia and Surface Dysgraphia in Adults and Children.* Lawrence Earlbaum.

Gentry, J. (1982). An analysis of developmental spelling in GNYS AT WRK. *The Reading Teacher*, *35*, 192–200.

Jordan, M. I. (1986). Serial order: a parallel distributed processing approach. Technical Report 8604, University of California, Institute for Cognitive Science, San Diego.

Karmiloff-Smith, A. (1979). Micro- and macrodevelopmental changes in language aquisition and other representational systems. *Cognitive Science*, *3*, 91–118.

Karmiloff-Smith, A. (1986). From meta-processes to concious access: evidence from children's metalinguistic and repair data. *Cognition*, *23*, 95–147.

Karmiloff-Smith, A. (1988). The child is a theoretician, not an inductivist. *Mind and Language*, *3*(3), 183–195.

Karmiloff-Smith, A. (1990). Constraints on representational change: evidence from children's drawing. *Cognition*, *34*(2), 57–83.

Liberman, I., Shankweiler, D., Fischer, F., & Carter, B. (1974). Explicit syllable and phoneme segmentation in the young child.. *Journal of Experimental Child Psychology*, *18*, 201–212.

Morais, J., Cary, L., Alegria, J., & Bertelson, P. (1979). Does awareness of speech as a sequence of phones arise spontaneously?. *Cognition*, *7*, 323–331.

Pinker, S. & Prince, A. (1988). On language and connectionism: analysis of a parallel distributed processing model of language aquisition. *Cognition*, *28*, 73–194.

Rumelhart, D. E. & McClelland, J. L. (1982). An interactive activation model of context effects in letter perception: part 2. the contextual enhancement effect and some tests and extensions to the model. *Psychological Review*, *89*(1), 60–94.

Rumelhart, D. E. & McClelland, J. L.(Eds.) . (1986). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition.* MIT Press, Cambridge, MA.

Seidenberg, M. & McClelland, J. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, *96*(4), 523–568.

Seymour, P. & Elder, L. (1986). Beginning reading without phonology. *Cognitive Neuropsychology*, *3*(1), 1–36.

Stanovich, K., Cunningham, A., & Cramer, B. (1984). Assessing phonological awareness in kindergarten children: issues of task comparability. *Journal of Experimental Child Psychology*, *38*, 175–190.

Stuart, M. & Coltheart, M. (1988). Does reading develop in a sequence of stages?. *Cognition*, *30*, 139–181.

Stuart, M. (1990). Factors influencing word recognition in pre-reading children. *British Journal of Psychology*, *81*, 135–146.

Thompson, G. & Fletcher-Flinn, C. (1991). A theory of knowledge sources and procedures for reading acquisition. In Thompson, G., Tunmer, W., & Nicholson, T.(Eds.) , *Reading Acquisition Processes.* Clevedon, UK: Multilingual Matters, Victoria University, Wellington, New Zealand.