UNIVERSITÀ DEGLI STUDI DI TRENTO
Dipartimento di Scienze della Cognizione e della Formazione

Scuola di Dottorato in Scienze Psicologiche e della Formazione

PhD Thesis

Cascaded and thresholded processing in visual word recognition: does the Dual Route Cascaded model require a threshold?

Supervisor:
Prof. Remo Job

Co-supervisor:
Prof. Claudio Mulatti
Prof. Max Coltheart

PhD Candidate:
Veronica Cembrani

Anno Accademico 2009-2010

CONTENENTS

1 GENERAL INTRODUCTION

The development of computational models during the lasts decades has revolutionized the scientific research in psychological sciences. The idea to model cognition is not new; more than forty years ago, in fact, Neisser (1967) provided a definition of cognitive psychology characterizing people as dynamic information-processing systems whose mental operations might be described in computational terms. Computational models are usually described as having many advantages over representing the theory about cognition as "*verbal model*" (Jacobs & Grainger, 1994). Any attempt to develop a computational model requires indeed completeness since a program will not run unless the theory is fully specified. Moreover, expressing any theory in computational terms immediately reveals many ways in which that theory may be incomplete or underspecified. Once the theory is complete and the program is executable, the adequacy of the theory can be tested by simulations. The comparison of the human behaviour and the behaviour of the computer program in carrying out the cognitive activity of interest is a straightforward and powerful way to test the validity of our scientific accounts. A theory has in fact to be sufficient, thus offering an explanation of all the relevant empirical phenomena. Mismatching between human behaviour and the behaviour of the computational model often revels ways in which the theory is incorrect thus suggesting how it can be reformulated in order to eliminate those mismatches. Other times the mismatches are so fundamental that modelling leads to theory refutation.

As in other cognitive fields, the interest in computational modelling is evident in psycholinguistic and, in particular, in the research on reading aloud. Reading, defined as the ability to generate a phonological code from print, has been largely studied over the last century in cognitive psychology as evidenced by numerous scientific papers and specialist journals, national and international conferences and founding of societies focusing on reading and reading impairments. One of the major aims of these researches has been to explain the complex cognitive processes involved in reading with the final goal of determining a complete theory of reading aloud and visual word recognition. Nevertheless, although these studies have resulted in a remarkable accumulation of knowledge and computational models of visual word recognition and reading aloud have been recently presented (e.g., Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Grainger & Jacobs, 1996; McClelland & Rumelhart, 1981; Perry, Ziegler, & Zorzi, 2007), the reading process is still not fully understood.

The aim of the current thesis is to expand the current understanding of visual word recognition in healthy skilled readers and to interpret this evidence within a computational account. We in fact believe that the cognitive processes of interest cannot be understood without a strong theoretical framework and that computational modelling is a highly valuable approach. The final goal of this activity will be to evaluate cascaded processing generally assumed in visual word recognition by referring to recent empirical data claiming the need of a different interpretation (i.e., thresholded processing).

This thesis focus on the Dual-Route Cascaded (DRC) model of reading aloud and visual word recognition (Coltheart et al., 2001). The DRC model has been chosen as the referential computational model in my thesis for several reasons. First, this thesis aims to evaluate cascaded processing, one of the main assumptions within the DRC framework. Second, the DRC model is not restricted to the simulation of English data in that an Italian version of the model has been recently developed (Mulatti, 2005; Mulatti & Job, 2003a; Mulatti & Job, 2004). Finally, there is a large agreement within the scientific community in recognizing the DRC model (one of) the most successful computational model(s) of reading, since "*the set of phenomena that the DRC model can simulate is much larger than the set that any other current computational model of reading aloud can simulate*" (Coltheart et al., 2001, p. 251).

1.1 Computational models of reading and visual word recognition

A computational model is defined as "*a computer program that is capable of performing the cognitive task of interest and does so by using exactly the same information-processing procedures as are specified in a theory of how people carry out this cognitive activity*" (Coltheart et al., 2001, p. 204). There are basically two approaches to generate computational models.

One of them is to develop connectionist models through a learning algorithm like the back propagation (e.g., Plaut, McClelland, Seidenberg, & Patterson, 1996). Computational models of this type are usually based on a network with three layers (input units, hidden units, output units) and a random value is initially assigned to the network connection weights. These values are then adjusted during the training in such a way to make the response for each stimulus more and more close to the correct response.

A possible disadvantage of this type of models is that it could be very difficult to discover the functional architecture of a trained network. In other words, it is rarely clear how the trained network has been structured by the learning algorithm and, thus, it is often not understandable how the model is actually working in performing the task it has learned.

Another approach consists in pre-specifying the functional architecture of the model, rather than relying on learning algorithms to do this. In this approach, even if the functional architecture is specified by the modeller, some form of learning algorithm may be used to set the strength of the connections between the pre-specified levels of the architecture.

The computational model that will be analyzed in this thesis – the DRC model – has been developed by applying this latest method. Only this approach will be considered in this dissertation.

1.1.1 Cascaded and thresholded processing

Many different mental operations are usually involved in cognitive activities, even in the most automatized. As a consequence, when one want to analyze performance in a task involving information processing, this is typically decomposed into a set of separated sub-processes. When modelling cognition, different levels of processing corresponding to the different mental operations are thus assumed and implemented in the system.

The functional architecture of highly successful computational models of reading (e.g., Coltheart et al., 2001; McClelland & Rumelhart, 1981; Perry et al., 2007) consists of different levels[1] of processing that get activated when a letter string is presented. In general, the levels in the reading system are domain-specific: each level has a specific function and different levels have different functions. Also, the levels are hierarchically organized, so that every level forms a representation of the input at a different level of abstraction. Each level assumed in the model usually consists of a number of processing units (or nodes), which accumulate information in the form of activation; units of the different levels communicates with several others (either at the same or at a different level of processing) and communication proceeds through the different levels via a spreading activation mechanism.

The notion of word perception as taking place in a hierarchical information-processing system isn't new and, also in the past, it has been advocated by several researchers interested in word perception (e.g., Adams, 1979; Estes, 1975; Johnston & McClelland, 1980; LaBerge & Samuels, 1974; McClelland, 1976). Regardless of it is today widely accepted that multiple levels of representation are involved in visual word recognition, the organization of these levels is otherwise controversial. In particular, it is still not clear how communication proceeds through the different

[1] The term *module* as referring to a domain-specific cognitive processing system is also used (e.g., Coltheart, 1999; Coltheart et al., 2001).

levels assumed in the reading system and thus, more generally, how information is processed in visual word recognition. In computational terms, the problem concerns the implementation of the activation function spreading through the different levels of processing assumed in the model.

Specifically, two major accounts have been proposed in this context. These accounts are schematically represented in Figure 1.

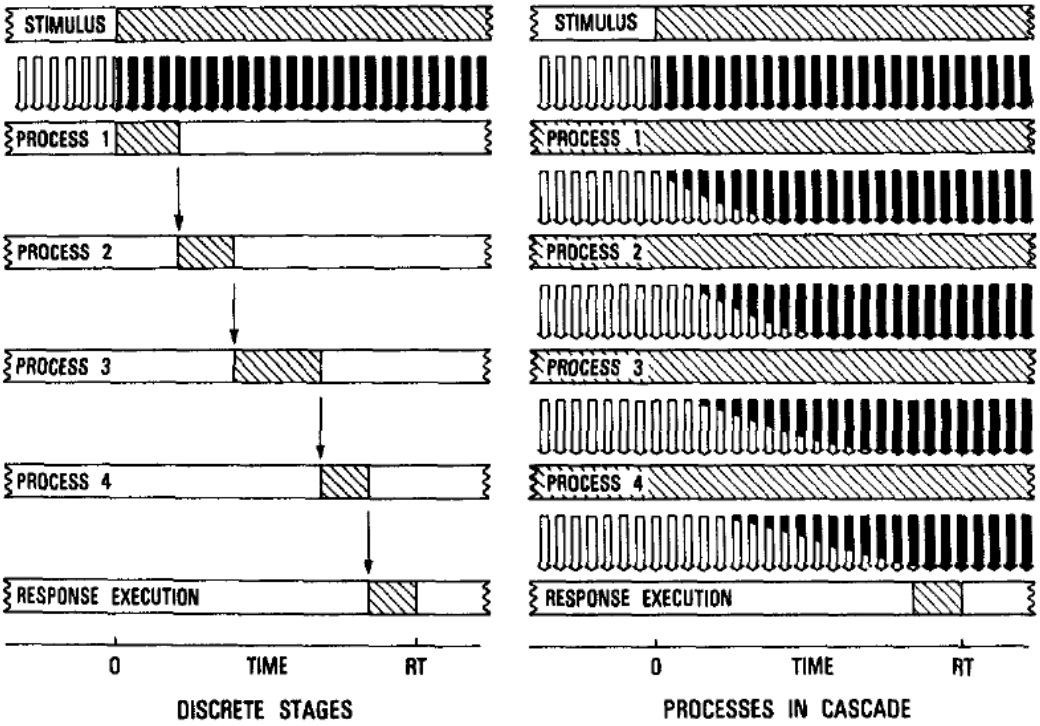

Figure 1. The events that occur between the presentation of a stimulus and the execution of a response, according to the discrete stage model and to the cascaded model. Arrows represent the transfer of information from one level to the subsequent one and shading is used to indicate when a process is at work; the blackening of the arrows indicates the degree to which the signals represented by the arrows reflect the stimulus in input (McClelland, 1979, p. 290, Figure 1).

One hypothesis is that performance may be represented by a model in which sub-processes are identified as successive temporal stages, each of which occupies a separate interval of time. This idea dates back to Sternberg (1969), who proposed that many mental processes occur in discrete series, one beginning when another ends. In models that assume this type of processing (e.g., Morton, 1969), activation does not propagate forward through the levels until processing within a level has reached some threshold (i.e., thresholded processing). Usually, activation is only passed on to the later stages after processing is ended in the earlier level. This means that the processing going on in any level does not begin to affect subsequent levels at an early point in processing; conversely, none of the processes can begin until the preceding process is completed.

An alternative to thresholded processing is that activation propagates in a cascaded fashion in the system. McClelland (1976; 1979) proposed that mental processes are cascaded, thus overlapping in time. According to this account, a process does not begin when the previous one is ended; on the contrary, information is transferred between the different levels of processing all of the time. As a consequence, each sub-process in the system is continuously active and its output always available for processing in the subsequent levels; the activation of a particular processing unit in a certain level of the network would thus increase with time (up to some asymptotic value), depending on the strength of the input to it. In models that operate by cascaded processing (e.g., Coltheart et al., 2001; McClelland & Rumelhart, 1981) there is not threshold within the levels and, as soon as activation accumulates in a level, it spreads immediately to the adjacent one. This means that activation accumulates downstream in the system, without waiting for processing completion in the early levels.

The different activation propagation modalities described above are alternative hypothesis and, usually, models implement either cascaded processing or thresholded processing. Despite cascaded processing is assumed in the most computational accounts of visible language, which approach modelling cognition better fits the empirical evidence is not yet fully determined in theories of visual word recognition.

In the following paragraphs I will focus on two models of visual word recognition, the first assuming thresholded processing (i.e., the *Logogen* model), the other assuming cascaded processing (i.e., the Interactive Activation model); both the accounts are central in the development of the Dual-Route Cascaded model of reading.

1.1.2 The *Logogen* model

The *logogen* model isn't a computational model of language processing, but rather a theory expressed by the box-and-arrow notation[2].

Morton (1969) based his theorizing on the concept of a mental lexicon, which he described as a system of knowledge about word forms. He referred to this level as *logogen* system. Another type of lexicon was assumed in his theory, which is a system of knowledge about word meaning

[2] This notation will be used to represent the computational models described later in this chapter: the different levels assumed in the theory are represented by different boxes, whereas the arrows between these boxes represent the connections assumed in the theory between the levels.

(i.e., a cognitive system). The *logogen* system is a set of elements called *logogens*, one for each word in the model's vocabulary. *Logogens* are evidence-collecting devices with threshold. Evidence is collected from visual or auditory input and when the amount of evidence collected by a word's *logogen* exceeds that *logogen*'s threshold, information about that word in the cognitive system (i.e., the word meaning) also becomes available as a response in the response buffer. The more frequent a word, the less evidence is needed to reach the threshold, because each *logogen* has a resting level of activation whose value is proportional to the frequency of occurrence in the language of that *logogen*'s word.

An input *logogen* system, responsible for word recognition, and an output *logogen* system, responsible for word production, are assumed in the model. The input *logogen* system consists of a visual input *logogen* system, responsible of written word recognition, and of an auditory input *logogen* system, responsible of spoken word recognition (see Morton, 1979). Similarly, the output *logogen* system consists of an output *logogen* system for speaking and of an output *logogen* system for writing (see Morton, 1980). Finally, grapheme-phoneme and acoustic-phoneme routes are assumed in the model (see Morton, 1980) in order to make it able to process nonwords (i.e., strings of letters without a meaning).

The different *logogen* systems assumed in the model proposed by Morton constitute the levels of processing assumed in the DRC model. However, whereas the *logogen* model theorized both spoken and written language, the DRC model is current applied to visual word recognition only. Importantly, the evolution of the *logogen* model has been entirely data driven[3]: complexities were added in order to explain the empirical results that a previous and simplest version of model could not explain. Hence, the complex form of the DRC model, which was inherited from the final version of the *logogen* model, is motivated by a series of empirical findings.

[3] Besides other empirical phenomena, the *logogen* model has been proposed to explain empirical data obtained in the repetition priming paradigm. Some studies (e.g., Clarke & Morton, 1983; Winnick & Daniel, 1970) showed that the cross-modal repetition priming occurs in two subsequent tasks such as picture naming and reading aloud only when the interval between the two tasks is very short. Furthermore, it has been shown that hearing a spoken word do prime the subsequent recognition of its printed form only with very short intervals between prime and target. These data suggest, from one hand, that the input and the output lexicons have to be separated and, from the other, that two separated systems, one for speaking and the other for writing, are required. Without a similar distinction a cross-modal priming as those described above would be in fact expected even with long inter-trial interval. Moreover, the distinction between the different systems is supported by a large deal of cognitive-neuropsychological data. For example, in the condition known as word-meaning deft, printed words can be understood but spoken words cannot even if hearing is adequate (Bramwell, 1897; Howard & Franklin, 1988), and the reverse holds for pure alexia, a condition in which spoken words can be recognized but printed words cannot, even though vision is adequate (Déjerine, 1982; Coltheart, 1998). This double-dissociation clearly suggests that two input lexicons, one for printed and one for spoken words, are needed. Furthermore, some people suffering of brain damage have an impairment of the ability to produce spoken words with relatively intact writing and spelling (Lhermitte & Derouesné, 1974), and other people have an impairment of writing and spelling with relatively intact ability to produce spoken words (Basso, Taborelli, & Vignolo, 1978). This double-dissociation clearly suggest a similar organization of the output lexicon, with an output lexicon producing spoken words and another producing written words. For a further treatment about the existence of different lexicons in the human language-processing system see Coltheart (2004).

1.1.3 The Interactive Activation model

The first computational model of reading is known as Interactive Activation (IA) model (McClelland & Rumelhart, 1981).

The IA model has been developed by its modellers with the main purpose to account for the word superiority effect; this effect is attributed to Reicher (1969), who showed that perception of a letter is facilitate when it is presented in the context of a word than when it is presented in a random sequence of letter[4].

The different mental computations involved in visual word recognition are represented in the IA model as involving three hierarchically organized purpose-specific levels of processing: the visual feature level, the letter level and the word level.

The general architecture of the IA model is represented in Figure 2.

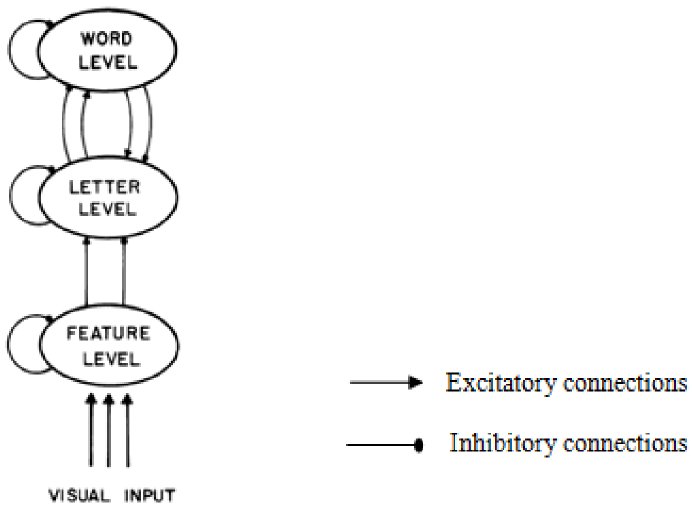

Figure 2. The processing system involved in visual word recognition (McClelland & Rumelhart, 1982, p. 379, Figure 2).

There is a strong similarity between the IA model and the *logogen* model. The IA model can be in fact considered "*a hierarchical, nonlinear, logogen model (...) with feedback between levels and inhibitory interactions among logogens at the same level*" (McClelland & Rumelhart, 1981, p. 388). The main difference between the *logogen* model and the IA model is that the units assumed in the latter model are not thresholded devices as the *logogens* assumed in the former. Instead, activation

[4] This effect is well-established in researches on visual word recognition. Before Reicher, previous researchers showed the word superiority effect in tachistoscopic presentation conditions (see Huey, 1908 and Neisser, 1967, for reviews). The problem of these studies, however, was that the effect was obtained during whole reports of all the letters presented and, since these reports were subjected to guessing biases and forgetting for longer stimuli, it wasn't clear whether the context in which a letter was presented influenced the process of perception itself rather than post-perceptual processes.

is assumed to spread through the different levels in a cascaded fashion. This means, for one hand, that the IA model is spatially parallel within the same level and, from the other, that it involves processes that operate simultaneously at several different levels. Crucially, perception is intended to be an interactive process in the IA model: there are bottom-up processing (i.e., feed-forward connections) and top-down processing (i.e., feedback connections) that work simultaneously and in conjunction, jointly determining what we perceive. Moreover, the communication in the system consists of both excitatory and inhibitory messages: excitatory messages increase the activation level of their recipients, while it is decreased by inhibitory messages. Furthermore, intra-level inhibitory loops are assumed in the model and represent a kind of lateral inhibition in which incompatible units at the same level compete each other.

In each level, for every relevant unit in the system an entity called node is assumed. Thus, there is a node for each word the model knows, and there is a node for each letter in each letter position within a four-letter string[5]. Since the nodes are organized into levels, there are word level nodes and letter level nodes. Each node has connections to a number of other nodes: if two nodes suggest each other's existence, then the connections between them are excitatory; on the opposite, if the two nodes are inconsistent with one another, then the relationship is inhibitory. The amount of excitation and inhibition that each node sends to the others is proportional to its activation, i.e. more active nodes send more activation than less active nodes. Each node has a momentary activation value and is said to be active when this value is positive. In the absence of inputs from other nodes, all the nodes are assumed to decay back to an inactive state (i.e., to an activation value at or below zero). The resting value differs from node to node and is determined by the frequency of activation of the node over the long term.

Connections may occur within levels or between adjacent levels, but there are no connections between non-adjacent levels. Connections within the word level are mutually inhibitory, since only one word can occur at any one place at any one time. Connections within the letter level are similarly organized. Connections between the word level and the letter level may be either inhibitory or excitatory, depending on whether the letter is a part of the word in that specific letter position.

Consider now what happens when an input reaches the system. When a stimulus is presented certain visual features are extracted and a set of feature inputs is thus made available to the system. The visual features are assumed to be binary in the model: thus, either the presence or the absence of a particular feature can be detected. The activation immediately spreads to the letter level: letter nodes that contain the extracted features are activated whereas letter nodes that do not contain those

[5] The computational version of the IA model implements four-letter strings only.

features are inhibited. The letter nodes, in turn, begin to send activation to those words that contain that letter in that particular position; also, each letter node inhibits those nodes representing words that do not contain that letter in that particular position. Within the letter level, each node representing a letter in a particular position inhibits all the nodes representing different letters in the same position. As the word level nodes become active, each word node starts to compete with all the others. In addition, each active word node sends feedback activation to the letter nodes. If the input features are similar to the features that form a particular set of letters and those letters are consistent with the letters forming the word that has been activated, the positive feedback in the system will converge on the appropriate set of letters and on the appropriate word. Otherwise, the active units inhibit each other and it might be that no single set of letters or single word will obtain enough activation to dominate the others.

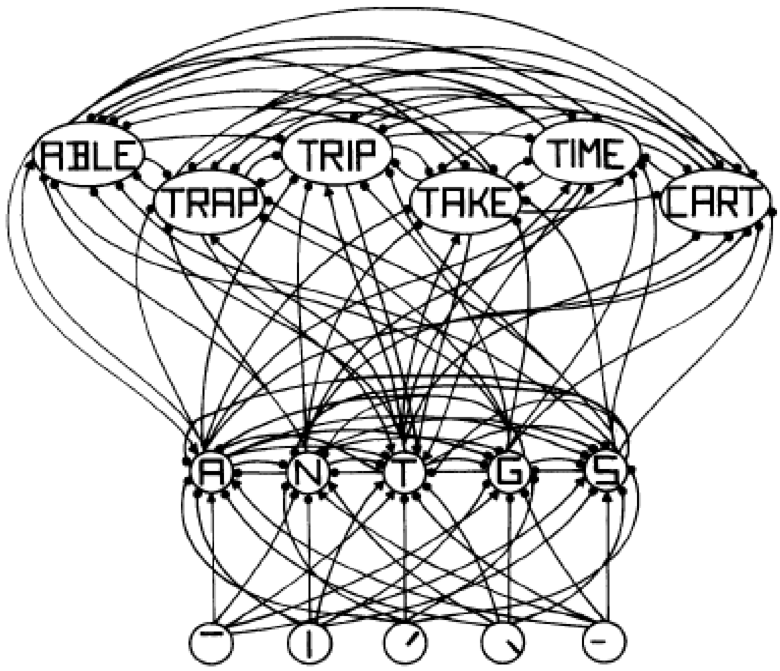

Figure 3. A few of the neighbours of the node for the letter “T” in the first position in a word and their interconnections (McClelland & Rumelhart, 1981, p. 380, Figure 3).

Computational simulations performed by McClelland and Rumelhart (1981) clearly showed that the IA model is able to account for the word superiority effect (Reicher, 1969). In particular, this effect is due to the interactive-activation between the word and the letter levels assumed in the model. Once a set of features is made available to the system and the letters consistent with those features have been activated, activation spreads to the word level. When the stimulus in input is a word, a node representing that word is activated and, in turn, it will send activation back to the letter nodes. Letter nodes that are consistent with the activated word will be excited, whereas letter nodes that are inconsistent with that word will be inhibited. This means that the feedback from the word level assists target letter recognition by contributing activation to the appropriate nodes and inhibition to the inconsistent nodes at the letter level. On the contrary, when the stimulus in input is not a word, there is not feedback from the word level; hence, the letter nodes receive activation from the feature level only and the identification of the correct letter will be slower and less accurate.

Despite the origin of the IA model, the word superiority effect is not the only evidence that it is able to explain. On the contrary, this model correctly accounts for various findings of several experiments in word perception (see McClelland & Rumelhart, 1981, for a detailed treatment). To date, the IA model has not been refuted. This observation has been critical for the development of the subsequent DRC model. In science, a new theory should account for all the crucial effects explained by the previous generation of the same theory or by other competitive theories, plus some new empirical data. This may be easily achieved by including a previous theory in a new model. According to the principle of nested incremental modelling, in fact, "*a new model should be related or include at least its own direct predecessor*" (Jacobs & Grainger, 1994, p. 1329). Following this principle, a generalization of the IA model has been included by Coltheart et al. (2001) in their Dual-Route Cascaded model of reading.

1.2 The Dual-Route Cascaded model

The Dual-Route Cascaded model (Coltheart et al., 2001) is based on traditional dual-route theories of reading aloud (e.g., Morton & Patterson, 1980) and, as said, has evolved from Morton's (1969) *logogen* model and McClelland and Rumelhart's (1981) interactive activation (IA) model of performance in perception task.

The DRC model adopts the architecture of the *logogen* model but avoids its theoretical commitment to the idea that the system operates according to the *logogen* principle; in other words, the assumption that the mental lexicon is composed of information-gathering devices with threshold is refuted. Instead, the central assumption of the DRC model is that activation propagates in a cascaded fashion through the different levels of processing; also, as in the IA model, interactive activation is assumed so that, as soon as there is activation in an early stage, it flows on to later stages and, having activated a subsequent level, it also feeds back to previous ones.

The overall architecture of the DRC model is represented in Figure 4.

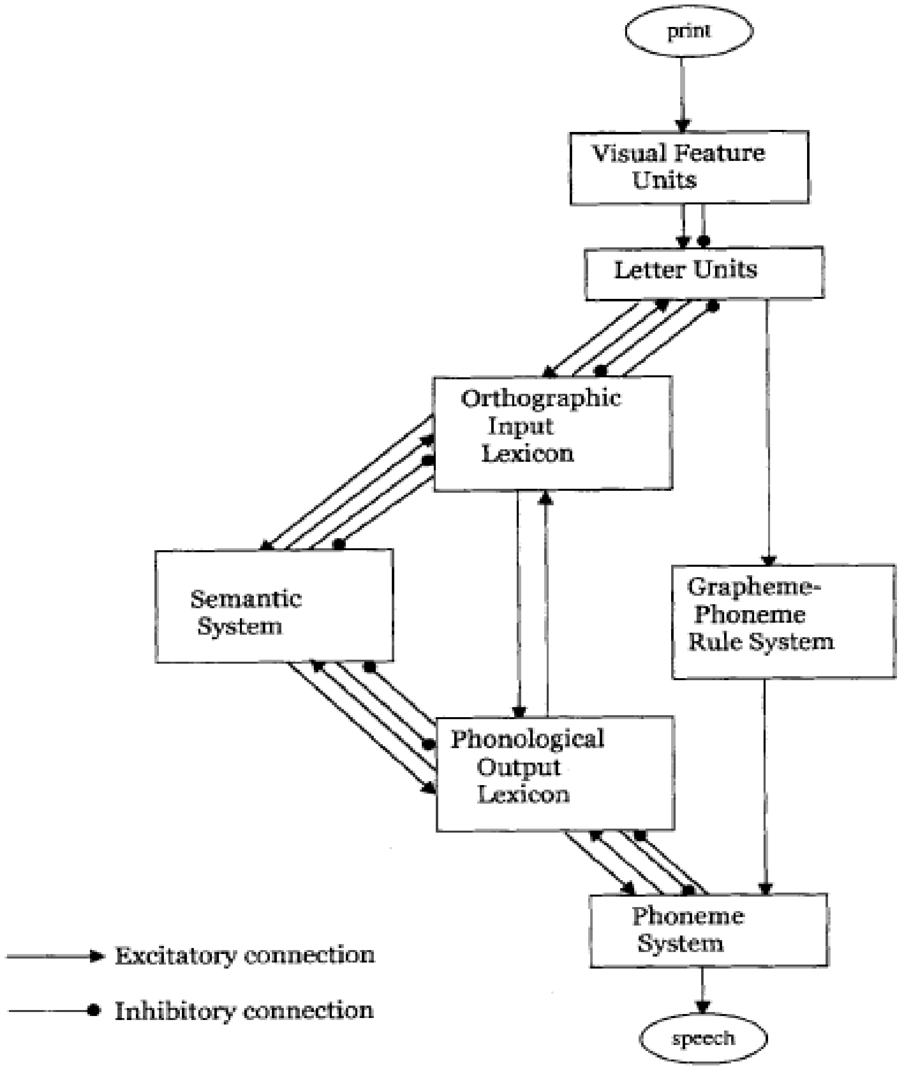

Figure 4. The Dual-Route Cascaded model of visual word recognition and reading aloud. (Coltheart et al., 2001, p. 214, Figure 7).

Two procedures to generate the phonology of a letter string are assumed in the model, the lexical route and the non-lexical route. Each route is composed of a number of levels containing set of units interacting each other through excitation (i.e., the activation of a unit contributes to the activation of other units) or inhibition (i.e., the activation of a unit makes more difficult the activation of other units). The units represent the smallest individual symbolic parts of the model, such as words and letters. Within the same level units interact each other only through lateral inhibition. Adjacent levels communicate fully in both directions in both excitatory and inhibitory ways[6].

[6] There are three exceptions here: first, the connection between the feature and the letter levels is only in one direction (i.e., from features to letters) as in the IA model; second, the communication between the orthographic and the phonological lexicons is only excitatory; third, the non-lexical route is feed-forward only.

The first two levels assumed in the model (i.e., the feature level and the letter level) are common to the two routes. In the model, feature analysis is spatially parallel and feeds forward to the letter level which processing is again spatially parallel. This portion of the DRC model is a generalization of the IA model. The only difference between the two models is that whereas the IA model applied to four-letters words only, its DRC version applies to words of any length up to eight letters. Hence, the visual feature level consists of eight different subsets representing the eight possible input positions; each subset consists of individual features that are set to on or off depending on whether that specific visual property is part of the letter in that specific position. The letter level also consists of eight different subsets and each subset contains units representing the entire set of letters that can occur (i.e., 26 letters in the English alphabet) plus one unit for the blank letter. Lateral inhibition occurs at this level, within but not between each of the eight subsets. The later inhibition will assure that only one letter occurs at any one place at any one time. The output from the letter level feeds both the lexical and the non-lexical routes.

The lexical route consists of two interconnected lexicons: the orthographic lexicon contains a single node (lexical entry) for each uniquely spelled word the model knows[7]; the phonological lexicon contains a single node for each uniquely sounding word the model knows.

The non-lexical route consists of a grapheme-phoneme correspondence (i.e., GPC) rules system and works serially along the string of letters. The GPC rules have been chosen on purely statistical grounds (i.e., for any grapheme, the phoneme assigned to it was the phoneme most commonly associated with that grapheme in the set of English monosyllables that contain that grapheme). Single-letter, multi-letter and context-sensitive rules are used for translating graphemes into phonemes.

The output from both these routes activates the phoneme system which is where the final pronunciation is produced. The phoneme units are similar organized to the letter units, except that each of the eight subset contains units for the phonemes that can occur (i.e., 43 in the English language) plus an unit for the blank phoneme. Pronunciation occurs when all the phonemes of the letter string have been activated to some criterion of satisfaction in the phoneme system. The DRC model operates over time units called cycles and the number of cycles it takes to reach criterion is considered a measure of the DRC model's response latencies.

Given the dual-route architecture, the phonological code of a string of letters visually presented can be generated in two different ways within the DRC model by employing the lexical or the non-lexical route.

---

[7] Only English mono-syllabic words are actually contained in the orthographic lexicon of the DRC model. These words are the 7,981 units of the CELEX database (Baayen, Piepenbrock, & van Rijn, 1993).

The pronunciation of the words in the orthographic lexicon is generated by the lexical route[8]; this is a parallel procedure that retrieves the whole-word phonology from stored lexical representation. When a word reaches the system, feature units activate the letter units (in parallel across letter position), which in turn activate words in the orthographic lexicon. The word units that have been activated activate the letter units via feedback and the phonological lexicon via feed-forward connections; finally, activation in the phonological lexicon feeds back to the orthographic lexicon and feeds forward to the phoneme system activating word's phonemes (in parallel across all phoneme positions), thus allowing the pronunciation of that word. A central feature of the lexical route is that units in the orthographic lexicon are frequency-sensitive: the activation of high-frequency words raises more quickly that the activation of low-frequency words. To achieve this effect, a constant value is associated with each unit in the lexicon. In languages with shallow orthographies, the lexical route is necessary to read irregular (or exception) words, i.e. words that disobey to the rules relating graphemes to their pronunciation[9].

The non-lexical route is a serial procedure that allows the model to read nonwords through the letter-by-letter conversion – from left to right – of each grapheme into the corresponding phoneme following language-specific correspondence rules. This route works as follow. Visual features and letter units are activated just as with the lexical route. Then the GPC route operates after a number of cycles. The set of rules is searched until an appropriate rule is found to convert the first letter to a phoneme and that phoneme unit in the phoneme system receives some activation. The next letter become then available to the GPC route[10] and the correct rule to translate that letter into the right phoneme is searched. Once all the letters are matched, pronunciation can occur.

Both the lexical and non-lexical routes are assumed to operate simultaneously – in parallel – on each stimulus. This means that when a letter string is presented to the system, activation from the feature units reaches the letter level and in turn both the orthographic lexicon and the GPC system. From one hand, cascaded processing from the orthographic lexicon eventually leads to a build-up of activation in the phoneme system (which also feeds back to the previous levels); at the same time, the GPC system contributes activation to the phoneme system. This feature allows the model to

[8] The lexical non-semantic route is described here. A lexical semantic route is also assumed in the model, but a semantic module is not yet implemented in its computational version.

[9] The distinction between regular and exception words is important in studying word-recognition in languages with inconsistent orthography like English. In languages with inconsistent orthography, the grapheme-to-phoneme mappings are quite irregular; on the contrary, languages with more transparent orthographies like Italian are characterized by consistent grapheme-to-phoneme mappings.

[10] A new letter becomes available to the GPC route after a constant number of cycles in the version of the DRC model originally described by Coltheart et al. (2001). In the last computational version of the model (i.e., the DRC 1.2), instead, the route moves on the next letter when the right-most phoneme set that was excited by the GPC route on the previous cycle contains a phoneme with an activation level that meets or exceeds a critical value.

account for effects due to the influence of the non-lexical procedure in word reading, such as the regularity effect (Seidenberg, Waters, Barnes, & Tanenhaus, 1984; Taraban & McClelland, 1987) or the position of the irregularity effect (Rastle & Coltheart, 1999), and the effects due to the influence of the lexical route on nonword reading, such as the pseudohomophone effect (McCann & Besner, 1987; Reynolds & Besner, 2005; Seidenberg, Peterson, MacDonald, & Plaut, 1996; Taft & Russell, 1992), the neighbourhood size (N) effect (McCann & Besner, 1987) or the position of the diverging letter effect (Mulatti, Peressotti, & Job, 2007).

A substantial amount of empirical data showed by skilled readers both in reading aloud and in lexical decision tasks, as well as a variety of behaviours exhibited by patients with various form of acquired dyslexia, are presently accounted for and correctly simulated by the DRC computational model (see Coltheart et al., 2001, for a more detailed discussion). Although subsequent works have revealed potential limits[11], to date it is perhaps the most successful computational model of reading today available in literature.

1.3 On the cascaded processing: some empirical data

The assumption of cascaded processing is central in the DRC model but also in many other frameworks of reading aloud; computational accounts of visible language processing are in fact almost invariable cascaded and often engaged in interactive-activation between the various levels of

[11] One of the major problems of the DRC model is, for example, the simulation of the effects depending on stimulus body (i.e., for a monosyllabic letter string, its body is the sequence of letters from its first vowel to the end) such as the consistency effect (e.g., Andrews & Scarratt, 1998; Glushko, 1979; Jared, 1997, 2002) or the body neighborhood (body-N) effect (e.g., Brown, 1987; Forster & Taft, 1994; Jared, McRae, & Seidenberg, 1990). The consistency effect is the following: body consistent stimuli (i.e., stimuli which body has the same pronunciation in all the words containing that body) are read faster than body inconsistent stimuli (i.e., stimuli which body has at least two different pronunciations in the set of words containing that body). The body-N effect is the following: stimuli comprising bodies that appear more frequently (e.g. *-eep*) are read aloud faster than stimuli comprising bodies that appear more rarely (*-eap*). The DRC fails to simulate consistency effects (see for example Jared, 2002) because its computational version does not include body representations. However, to overcome this problem, it is currently considered the possibility to add body-rime rules to the model's non-lexical route, as suggested by Patterson and Morton (1985). To date, both the consistency and the body-N effects are actually correctly simulated by the Connectionist Dual Process (CDP+) model (Perry et al., 2007), a model which overall architecture is very similar to that of the DRC model, as both rely on a lexical and on a non-lexical route to name the stimuli. In both the models, the lexical route is a symbolic localist interactive-activation network with activation propagating in a cascaded fashion, based on McClelland and Rumelhart's (1981) IA model. The main difference between the DRC model and the CDP+ model regards the way the non-lexical route works. Whereas the DRC's non-lexical route applies serially rules of correspondence between graphemes and phonemes, the CDP+'s non-lexical route is a two layers associative network (i.e. without hidden units) trained to learn the mapping between orthography and phonology and equipped with a serial graphemic parsing: during training, the CDP+ model non-lexical route acquires body-rime representations and the model is thus sensible to the effects depending on stimulus' bodies proprieties. Similarly, the consistency effect is simulated by parallel distributed processing (PDP) models (e.g., Harm & Seidenberg, 1999; Plaut, et al., 1996; Seidenberg & McClelland, 1989) that use learning algorithm to discover the relationship between spelling and sound.

processing. This is mainly because a certain number of well-established empirical findings strongly support cascaded processing in visual word recognition.

Besides several data may be certainly relevant in this context, cascaded processing in reading is mostly implicated by a number of effects showing lexical influence when reading nonwords. Suppose that processing is thresholded and information is processed in serial discrete stages in the reading system; if so, when a nonword is presented in a reading task, no entry in the orthographic lexicon will reach the threshold and hence no information will emerge from the lexicon. This wouldn't explain, for example, why nonwords inconsistent with real words (e.g., *heaf*; cf. *deaf*) yield longer reading latencies than nonwords (e.g., *hean*) that are not inconsistent with real words (Glushko, 1979).

In addition, such a thresholded processing is incompatible with numerous findings showing that the lexical route also influences the computation of phonology of the nonwords that have to be read (e.g., Rosson, 1983). This is evident, for example, in the so called pseudohomophone effect (e.g., McCann & Besner, 1987); this effect consists in pseudohomophone nonwords (i.e., nonwords which pronunciation matched the pronunciation of an existing word; e.g., *traxs*; cf. *tracks*) being read faster than control non-pseudohomophone nonwords (e.g., *prax*).

Further and strong evidence in favour of cascaded processing emerges from findings showing the effects of the orthographic neighbourhood size (N) in reading. The index N is defined as the number of words that can be created from a string of letters by replacing a single letter at a time (Coltheart, Davelaar, Jonasson, & Besner, 1977). For example, orthographic neighbours of *word* are *cord, ford, lord, ward, wood, wore, work, worm, worn.* Empirical evidence shows that the size of the orthographic neighbourhood influences reading latencies (e.g., Andrews, 1997, 1989; McCann & Besner, 1987; Job, Peressotti, & Cusinato, 1998; Peereman & Content, 1995). In skilled readers, words with many neighbours are read aloud more quickly than those with few or no neighbours; moreover, it has been largely showed that nonword reading is facilitate by a large N. The principal explanation of these effects is that they arise from activation within the lexical route (e.g., Andrews, 1997; Coltheart et al., 1977; Coltheart et al., 2001). Hence, the effects due to the orthographic neighbourhood in nonword reading can be explained only assuming cascaded activation in the reading system. More specifically, the N effect clearly suggests that the orthographic neighbours of a visually presented letter string are sufficiently activated to influence the computation of the phonology of the nonword that has to be read. Clearly, this activation would be completely prevented by assuming a threshold between the different levels of processing.

Despite a substantial amount of data supporting cascaded processing clearly exists, empirical results contrasting this assumption have been recently documented. These data have been principally obtained in reading aloud experiments examining the effect of the psycholinguistic variable of interest when factorially combined with the manipulation of stimulus quality in the task.

1.3.1 Factorial manipulations of variables: focusing on stimulus quality

Factorial experiments in which a factor affecting the rate of processing (e.g., stimulus quality) is varied in conjunction with another factor (e.g., word frequency, semantic priming, repetition priming, etc.) in reading tasks have been used in the last decades to evaluate different non-computational accounts of visual word recognition (e.g., Balota & Abrams, 1995; Besner & Smith, 1992; Besner & Swan, 1982; Borowsky & Besner, 1993; Meyer, Schvaneveldt, & Ruddy, 1975; Plourde & Besner, 1997; Stolz & Neely, 1995). Very recently, a certain number of works used this same approach to test the validity of computational models of reading (e.g., Besner & O'Malley, 2009; Besner, O'Malley, & Robidoux, 2010; Besner & Roberts, 2003; Blais & Besner, 2007; O'Malley & Besner, 2008; O'Malley, Reynolds, & Besner, 2007; Reynolds & Besner, 2004). To date, the most of these studies directly focus on the Dual-Route Cascaded framework.

When only a single factor is manipulated in experiments, the results can often be explained in a variety of different ways. In other words, many competing explanations – and several different (computational) models – may be equally accurate in explaining a main effect. However, when two factors are jointly manipulated, the data pattern is much more complex and it might help to falsify some of the various accounts. In particular, the manipulation of two different factors sometimes produces additive effects and sometimes produces interactions (i.e., either underadditive or overadditive effects) on response latencies. Hence, factorial manipulation of different variables is considered an useful toll in testing the validity of theoretical and computational accounts and a powerful investigation to distinguish between competitive theories.

Factorial manipulations in reading studies may consist in the psycholinguistic variable of interest being manipulated together with a factor affecting the rate of processing. Usually, the latter factor is the stimulus quality (or SQ), a perceptive variable affecting early processing in visual word recognition. A common technique to manipulate stimulus quality in experiments is by reducing the contrast between the visually presented stimulus and the background (e.g., Borowsky & Besner, 1993)[12]. The typical finding obtained in these tasks is that reading latencies increase for degraded stimuli. Importantly, stimulus quality is considered an useful second manipulation for testing the

[12] All the studies presented in this thesis as well as the most of the studies in literature that will be cited used contrast to manipulate SQ. However, other techniques to reduce stimulus quality have also been documented. For example, SQ can be reduced by presenting low-pass filtered stimuli (as in Fiset, Arguin & Fiset, 2006) or by alternating a mask and the letter string (as in Yap & Balota, 2007).

validity of computational accounts of reading because it can be simulated in most computational models. Typically, degradation is implemented by modifying the connection weights at one or more levels; specifically to the DRC model, degradation is usually simulated by reducing the weights of the excitatory and inhibitory connections that regulate the communication from visual feature to letter units. This manipulation would reduce the rate of activation gain at the perceptual level and responses will be thus slower for degraded than for clear stimuli. Various simulations documented, in fact, that DRC model's responses are delayed as the connection weights between the feature and the letter levels are reduced, thus miming the effect of degradation obtained for humans.

Strong evidence against cascaded processing has been reported in a number of different studies using the factorial manipulations described above. In this dissertation, the most critical results will be discussed. In particular, this thesis will focus on the studies that have examined the effects of SQ when factorially combined with:

1. letter string length when reading nonwords aloud (Besner & Roberts, 2003);
2. orthographic neighbourhood size (N) when reading nonwords aloud (Reynolds & Besner, 2004);
3. word frequency when reading aloud (O'Malley & Besner, 2008);
4. lexicality when reading aloud (Besner & O'Malley, 2009; Besner et al, 2010; O'Malley & Besner, 2008).

These studies will be examined in details in the next chapters of this thesis. At this step it is sufficient to note that SQ and the other variable have been shown to exert additive effects on skilled readers' latencies in all the experiments reported above. In other words, the effects of the psycholinguistic variables manipulated in these studies (e.g., letter string length, orthographic neighbourhood size, word frequency, lexicality) have been shown to have the same amplitude for stimuli presented in a clear (i.e., non-degraded) condition and for stimuli presented in reduced contrast in the human performance.

Critically, interactions between the two factors are on the contrary simulated by the DRC model, i.e. the amplitude of the effect of letter string length/orthographic neighbourhood size/word frequency/lexicality is significantly different for clear and degraded stimuli in the DRC model's simulations. These interactions would be caused by the cascaded activation assumed in the model. In fact, since processing is cascaded, a change in the rate of activation in early processing units (due to stimulus degradation) will change the rate of activation downstream in the model. As a consequence, SQ will likely to interact with variables affecting the subsequent levels of processing assumed in the reading system like letter string length, orthographic neighbourhood size, word frequency and lexicality.

1.3.2 Introducing a threshold in the reading system

The experiments described above provide significant mismatches between the DRC model's performance and the human behaviour; critically, these mismatches have been interpreted as caused by the cascaded activation assumed in this framework. As a consequence, the previous data contrast not only the DRC model but rather the same idea that activation proceeds in a cascaded fashion in the reading system.

Additive effects between variables are easily explained within a thresholded framework by postulating that those variables affect different levels of processing in the system. Sternberg (1969), following Donders (1868-1869), noted that one can attempt to study the component processes implicated in performance by using reaction times data and that additional assumptions about their temporal relations can be made by observing the pattern of results produced by factorial manipulations; in particular, one can use experimental manipulations that are assumed to selectively influence specific levels of processing to study what levels are affected by other manipulations. Within discrete stage models assuming thresholded processing, the assumption that one experimental manipulation influences the duration of one level and another manipulation influences the duration of another level leads to the conclusion that the two factors will have additive effects on reaction times; on the contrary, factors that influence the duration of the same level will generally interact with one another. If we adopt cascaded models, however, this logic is only partially correct. In cascaded models, activation propagates through the levels continuously, i.e. activation reaches a subsequent level *before* processing in a previous level is ended. This means that the effects of a factor is not resolved within the level it affects; rather, the effect of a factor at an early level would influence processing downstream in the system. This implies, from one hand, that factors interacting with one another could well be influencing different processes in a cascaded system and, from the other, that cascaded activation between different levels of processing is not easily reconciled with evidence showing additive effects between factors affecting those levels. It follows that the easier way to explain additive effects of factors is by assuming a threshold between the levels of processing which these factors affect. Additive effects are hardly explained within a cascaded framework especially when that framework also assumes interactive activation between the different levels of processing, as in the DRC model. In these circumstances, in fact, the effect of experimental manipulations influencing a specific process in the system not only cascades to the subsequent levels through feed-forward connections, but it also feeds back to the previous ones.

When factorial manipulations of psycholinguistic variables and stimulus quality are considered within the Dual-Route Cascaded framework, the previsions are the following. Cascaded activation assumed in the model will cause a variable affecting the rate of processing in early units (e.g., SQ) to affect the model beyond the perceptual level (i.e., the effect of SQ is not resolved at an early stage but rather affects processing downstream in the system); in turn, interactive activation will determine factors affecting later processes to feed their effects back to previous levels, thus having an effect on earlier factors. As a consequence, SQ will likely to interact in the DRC model with variables affecting subsequent levels of processing assumed in the system. As said, the DRC model produces in fact interactions of SQ with letter string length, orthographic neighbourhood size, word frequency and lexicality, inconsistently with the empirical data. In order to eliminate the mismatches with the human performance, a reformulation of the DRC model has been pointed out (e.g., Besner & Roberts, 2003; Blais & Besner, 2007; O'Malley & Besner, 2008; Reynolds & Besner, 2004). Even if partially different interpretations have been proposed to explain the different findings, all these solutions generally agree on a critical point: thresholding the letter level rather than allowing it to cascade provides a simple way to allow the DRC model to fit the additive effects produced by human readers.

From a theoretical perspective, in fact, SQ is a perceptual variable influencing the recognizability of letters and degradation would thus not affect the model beyond the letter level; when the letter level is thresholded, the effect due to degradation would be resolved within early levels of processing. This would prevent interactions between variables affecting the perceptual level (e.g., SQ) and variables affecting the subsequent levels of the model (e.g., letter string length, neighbourhood size, word frequency and lexicality), thus allowing the DRC model to explain the additivities that have been documented. Furthermore, this proposal is not a merely theoretical account. Instead, simulation works confirm that changing the model in this way is successful in that the DRC model so modified correctly simulates the additive effects of SQ and 1) letter string length in nonwords reading 2) orthographic neighbourhood size in nonwords reading and 3) word frequency in reading, consistently with the empirical data (see Besner, Reynolds, & Chang, 2003)[13].

This solution assumes, however, that at least some processes in the reading system occur in discrete series, one beginning only when the previous ends; in other words, information processing

[13] The authors didn't attempt, however, to demonstrate whether implementing the letter level as thresholded would allow the DRC model to simulate all the effects that its current version does simulate. In adherence to the principle of nested modelling, instead, any novel account (or any modification of an old model) should be showed able to reproduce all the effects that its previous version was able to simulate.

in visual word recognition would be thresholded, at least at the level processing letters. Clearly, a whole change of the DRC model is intrinsic in this proposal and accepting this modification would more generally mean to reject the idea of cascaded processing per se.

1.4 Goals and outline of the thesis

The current thesis aims to investigate cascaded processing in visual word recognition by testing the predictions of the Dual-Route Cascaded model of reading aloud. Despite widespread acceptance of the idea that visual language processing is cascaded, there are circumstances in which such an account is not easy reconciled with the data produced by skilled readers. In particular, recent experiments involving factorial manipulations suggested that the information processing implicated in visual word recognition might be at least partially thresholded. Information processing will be evaluated by referring to these studies; more specifically, the discussion of the previous results will be supported by the presentation of new empirical data obtained either in Italian or in English reading aloud tasks as well as by DRC model’s simulations.

Specifically, the thesis is structured as follow. The studies reported in the first chapters (Chapter 2 to 5) focus on factorial manipulations in nonword reading; the main aim of these studies will be to define whether the experiments in which SQ is manipulated together with a second factor (e.g., letter string length and orthographic neighbourhood size) in nonword reading tasks can be explained by considering a variable reflecting the visual similarity between the different letters of the alphabet, namely the Total Letter Confusability. The experiments reported in the last chapters (Chapter 6 and 7) will instead focus on factorial manipulations in reading as a function of the type of stimuli presented in the task; in Chapter 6 the effects of factors affecting the recognisability of letters (e.g., SQ and Total Letter Confusability) will be analyzed when jointly manipulated with lexical factors (e.g., word frequency and lexicality); the effects due to list composition in degraded presentation will be directly assessed in Chapter 7. The implications of these findings for theories of visual word recognition and computational models of reading will be discussed.

2 LENGTH AND ORTHOGRAPHIC NEIGHBORHOOD SIZE IN NONWORD READING

In this chapter the joint effects of letter string length and neighbourhood size (N; Coltheart et al, 1997) will be analyzed when reading nonwords aloud. These effects are particularly relevant for the present purposes given their manipulation in multifactor experiments involving degradation. The aim of the study presented in this chapter is to analyze the joint effects of letter string length and N in nonword reading when the stimuli are presented in a non-degraded (i.e., clear) condition. These effects will be interpreted within the DRC framework that assumes two routes – a lexical and a non-lexical procedure – operating simultaneously on each stimulus. In this study we explored a prediction the model makes with respect to nonword reading and that directly follows from its dual-route architecture and cascaded processing: the orthographic neighbourhood size effect should increase as letter length increases. The results of the experiment are consistent with this prediction.

2.1 Introduction

The DRC model of visual word recognition and reading aloud appeals to two procedures to generate the phonology of a letter string: the lexical route and the non-lexical route. The lexical route is a parallel procedure that retrieves the whole-word phonology from stored lexical representation and allows the model to read irregular words. The non-lexical route converts serially, letter by letter (from left to right) each grapheme into the corresponding phoneme following language-specific correspondence rules; this routine is necessary for nonword reading.

Regardless the characteristics of the stimulus (i.e. whether it is a regular or an irregular word, or a nonword) both the procedures are hired upon stimulus presentation. Also, both the procedures are assumed to work simultaneously – in parallel – on the stimulus.

The N effect in nonword reading (e.g., McCann & Besner, 1987) is of specific interest for the current purposes. The effect is the following: nonword reading times decrease as the number of its orthographic neighbors increases. Within the DRC framework, the N effect is accounted for by postulating cascaded processing along connections that allows nonwords to activate orthographically similar words in the orthographic lexicon. This activation spreads to the corresponding phonological representations in the phonological lexicon, which in turn send activation to the phonemic units in the phonemic buffer. Since the orthographic neighbours of a nonword and a nonword itself usually share many phonemes, phonemic activation generated by the lexical routine, paired with correct nonlexical processing, positively contributes to the assembling of the nonword phonology; in other words, the phonological representations activated by the orthographic neighbours prime the phonemes in the phoneme system (Coltheart et al., 2001; but see Reynolds & Besner 2002)[14]. Thus, according to the DRC model, the N effect in nonword reading is an effect due to the lexical route, which is a procedure that works in parallel on the stimulus. However, the DRC relies on the non-lexical routine to read nonwords, which is a procedure that works serially. This leads to a prediction: the size of the effect of N should increase as the length (i.e., number of letters) of the nonword increases. As the number of letters in a nonword increases, the time required to the non-lexical route to reach the last rightmost letter and activate the last phoneme increases as well. The longer the increase of time, the longer the lexical route works on the stimulus. Thus, the longer is the increase of time, the stronger the lexical route primes the neighbour's phonemes and facilitates nonword reading.

In the experiment reported below we tested this prediction. To this end we collected data from Italian skilled readers presenting short and long nonwords without orthographic neighbours and short and long nonwords with one or more orthographic neighbours. We also performed a simulation with the Italian version of the DRC model (Mulatti, 2005; Mulatti & Job, 2003a). We expected both humans and the DRC model to show: a) a main effect of neighbourhood size, b) a main effect of letter length, and c) an interaction between the two factors such that items with neighbours show a smaller length effect with respect to the items without neighbours.

2.2 Method

Participants. Eighteen students at the Università degli Studi di Padova who had Italian as their first language and normal or corrected-to-normal vision participated as part of their courses requirement.

[14] As will be discussed in Chapter 5, another account explaining the effect of N because of the interactive activation between the orthographic lexicon and the letter level has been proposed within the DRC framework. However, it is important to note that the prediction of the DRC model tested in the present study does not change regardless of the view adopted to explain the effect of N since both accounts assume that it arises within the lexical route.

Design. The experiment consisted of a 2 X 2 within-subject design with N (zero-N vs. one-or-more N) and Length (short vs. long items) as factors.

Materials. A total of 112 orthographically legal disyllabic nonwords were used in an Italian reading aloud task (these stimuli can be seen in the Appendix, section A). These consisted of 56 nonwords that had no neighbours and 56 nonwords that had one or more neighbours. Within each group of nonwords, there were 28 stimuli at each of two levels of length (short vs. long). Items that were 5 letters in length represented the short length condition; items that were 7 letters long represented the long length condition. Short and long nonwords were balanced with respect to the initial phoneme. In addition, for nonwords with neighbours (i.e. nonwords derived by changing one letter of an Italian word, provided the initial phoneme of that word remained intact), short and long items were balanced in terms of baseword frequency (3.7 vs. 3.8 occurrences per million, $t<1$), baseword neighbourhood size (2.2 vs. 2.1, $t<1$), nonword neighbourhood size (1.2 vs. 1.2, $t<1$; range: 1-3), nonword neighbourhood frequency (3.6 vs. 3.2, $t<1$) and the position of the letter changed (3.4 vs. 3.4, $t<1$; see Mulatti et al., 2007).

Apparatus. The experiment took place in a sound attenuated and dim lit room. Stimuli presentation and data recording were controlled by software developed in E-prime and running on a personal computer. Stimuli were presented centrally in black lower-case letters on a white background. The display was synchronized with the screen refresh cycle. Subjects' naming responses were detected via a microphone connected to a voice-key. Participants sat in front of the computer screen and the microphone was placed directly in front of but slightly below the subjects' face, so as not to obstruct screen view. Response latency was timed from stimulus onset to voice key activation, which also terminated the display.

Procedure. Participants were tested individually. They were instructed to read each letter string aloud as quickly and accurately as possible; they were informed that a letter string would be a pronounceable nonword. Each trial began with a fixation point (+) presented for 500 ms; then, the display went blank for 100 ms. Immediately after the stimulus appeared and remained on the screen until a response was registered by the voice-key or 3 sec elapsed. The inter-trial-interval was set to 2 sec. Stimuli were presented in six different pseudorandom orders across participants. A practice session preceded the experimental session and consisted of 12 items presented at each subject in a random order. The experimenter coded the pronunciation error as triggering (i.e. voice key failure), lexicalization and articulation fault.

2.3 Results

Articulation errors (11%) and apparatus failures (4.2%) were excluded from the analysis of reaction times (RTs); also, apparatus failures were excluded from the analysis of accuracy. Correct RTs were submitted to the Van Selst and Jolicoeur (1994) trimming procedure[15], which excluded an additional 1.3% of the data. Mean naming latencies and percentages of accuracy scores – according to conditions – are reported in Table 1.

	Orthographic Neighborhood					
	Zero-N		One-or-more-N		Diff	
Length	RTs	%E	RTs	%E	RTs	%E
Long	696	15	645	10	51	5
Short	637	13	626	8	11	5
Diff.	59	2	19	2		

Table 1. Mean reaction times (RTs) and percentages of error (E%) according to conditions.

ANOVAs with N (zero-N vs. one-or-more-N) and Length (short vs. long items) as repeated factors for the participant analysis (F1) and as independent factors for the item analysis (F2) were conducted on RTs and accuracy.

RTs. Analysis showed a main effect of N, $F1(1, 17) = 20$, $MSE = 830$, $p < .001$, $F2(1, 108) = 13$, $MSE = 3118$, $p < .001$, a main effect of Length, $F1(1, 17) = 16$, $MSE = 1688$, $p < .005$, $F2(1, 108) = 16$, $MSE = 3118$, $p < .001$, and, crucially, a significant interaction, $F1(1, 17) = 9$, $MSE = 746$, $p = .007$, $F2(1, 108) = 4$, $MSE = 3118$, $p < .05$, due to the fact that – as predicted by the DRC model – the size of the length effect is smaller for the items with neighbours with respect to the items without neighbours.

Accuracy. Analysis showed a main effect of N, $F1(1, 17) = 6.1$, $MSE = .006$, $p < .05$, $F2(1, 108) = 4$, $MSE = .014$, $p < .05$, whereas neither the effect of Length, $F1(1,17) = 1.7$, $p > .2$, $F2<1$, nor the interaction, $Fs < 1$, proved significant.

[15] Van Selst and Jolicoeur (1994) proposed a recursive data trimming procedure in which the criterion cut-off for outliers removal is established by the sample size in each condition for each subject. This method has the advantage to avoid problems due to sample size on outliers elimination.

2.4 Simulation

The Italian version of the DRC model (Mulatti & Job, 2003a) resorts to a vocabulary of Italian monosyllabic and paroxytone disyllabic words (i.e., words stressed on the second last syllable) since both are pronounced without reference to supra-segmental information. The architecture of the DRC and the parameter set governing lexical and non-lexical processing are those of the English version (Coltheart et al., 2001). There are 6,382 units in the orthographic input lexicon, and 6,372 units in the phonological output lexicon. As in the English version, the non-lexical route uses single-letter (e.g., *p* to /p/), multi-letter (e.g., *ch* to /k/), and context-sensitive (e.g., *c[i]* to /tʃ/) rules for translating graphemes into phonemes. The model correctly pronounces the whole set of items in its lexicon, and simulates the regularity effect observed with loan words (Mulatti, 2005; Mulatti & Job, 2003b; Schereer, 1987; Ziegler, Perry, & Coltheart, 2000) and the effect due to the position of the diverging letter in nonword reading (Mulatti et al., 2007).

The set of nonwords used with the participants was run through the Italian version of the DRC model. We chose to use the parameter set that allows the model to correctly simulate the neighbourhood size effects in reading aloud (see Coltheart et al., 2001, p. 224). The model did not make any error. Mean cycles to criterion are reported in Table 2.

	Orthographic Neighborhood		
	Zero-N	One-or-more-N	Diff.
Length	Cycles	Cycles	Cycles
Long	187	144	43
Short	154	132	22
Diff.	33	12	

Table 2. Mean cycles according to conditions.

An ANOVA with N and Length as independent factors was conducted on cycles. The DRC behaviour mimed that of humans. Analysis showed a main effect of N, $F(1, 108) = 94$, $p < .001$, a main effect of Length, $F(1, 108) = 45$, $p < .001$, and a significant interaction, $F(1, 108) = 10$, $p < .001$, imputable to the size of the length effect being smaller for the items with neighbours than for the items without neighbours.

2.5 Discussion

The experiment evidenced three different results, all correctly simulated by the DRC model.

First, nonwords with neighbours are read faster than nonwords without neighbours. In the literature, this result is referred to as the N size effect (e.g., McCann & Besner, 1987; see also Arduino & Burani, 2004). As mentioned in the introduction, within the DRC framework the N effect arises from activation in the orthographic lexicon that feeds forward to the phonological lexicon and primes the phonemes in the phoneme system.

Second, we found a length effect: short nonwords are read aloud faster than long nonwords. This result is consistent with that of Weekes (1997), who presented his participants with words and nonwords of three, four, five, and six letters for reading aloud. He found a main effect of letter length, a main effect of lexicality (i.e., words read faster than nonwords), and a significant interaction due to the fact that whereas nonwords showed a length effect, words showed no length effect (however, see Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004). Within the dual route framework, the lexicality by length interaction receives the following explanation: nonwords are assembled from letters one by one, hence causing a length effect, whereas words are retrieved from the lexicon as a whole, through a parallel process, hence preventing/attenuating serial effects.

Third, the size of the length effect depends upon the size of N: nonwords without neighbours exhibit a stronger length effect than nonwords with one or more neighbours. Within the dual route theory, this interaction is easily explained. If a nonword has one or more orthographic neighbours, their orthographic and phonological representations receive activation from the stimulus. Activation grows over time. Nonlexical processing proceeds letter by letter, and therefore longer nonwords require more time to be assembled. Thus, while processing longer nonwords, the activation in the lexical route grows for a longer interval of time reaching higher levels. Since the activation of the lexical route positively contribute to the assembling of the nonword's phonology, longer nonwords are more facilitated by their neighbouring words than shorter nonwords.

To conclude, a prediction of the DRC model has been tested through empirical investigation. The results of our experiment are consistent with this prediction and are all correctly simulated by the Italian version of the DRC model. The interaction between the letter string length and the orthographic neighbourhood size (N) that has been obtained is particularly relevant for the definition of a computational model of visual word recognition and reading aloud. This interaction strongly supports a cascaded model with a dual route architecture, comprising a route working in parallel and another working serially, as assumed in the DRC framework.

3 TOTAL LETTER CONFUSABILITY IN DEGRADED NONWORD READING

In this chapter the studies providing strong evidence against cascaded processing in nonword reading will be considered. As pointed out in the introduction, these experiments typically consist of a factorial manipulation involving stimulus degradation. A new variable that might play a role when stimuli are degraded – the Total Letter Confusability (TLC) – will be introduced. Since the letters comprising the stimulus in input are hardly identified when stimuli are degraded, the visual similarity between the different letters might influence letter identification. In fact, some letters in the alphabet are perceptively similar to others letters (e.g., E and F) whereas other letters are not (e.g., Z and J); similar letters might be thus confused more likely than other less-similar letters when the string is presented in reduced contrast. A measure of letter similarity – or letter confusability – could hence be a relevant factor to consider in researches analyzing the effects of stimulus degradation on speeded nonword reading, since one might expect more similar/confusable letters to suffer more when degraded than less similar/confusable letters.

3.1 Introduction

Any pair of letters has a visual similarity that can be defined by the number of features that the two letters have in common. The idea that letters are identified by their constituent part – their features – is not new and it has been proposed over 50 years ago by Selfridge (1959); in his model, the Pandemonium, letter identification was achieved by hierarchically organized layers of features and letter detectors. More recently, numerous researches provided convincing evidence in support of this account (see Grainger, Rey, & Dufau, 2008, for a review).

Letter similarity is sometimes referred to as letter confusability. The greater the visual similarity between two letters is, in fact, the more likely will observers be to confuse one of these letters with the other. Measures of letter confusability are empirically generated. Typically, isolated letters are presented in data-limited conditions (e.g., brief exposure and/or low contrast and/or masking) and participants are asked to report the presented letter. Error rates (e.g., reporting F when

E was presented) will give a measure of letter confusability between pairs of letters (e.g., Gilmore, Hersh, Caramazza, & Griffin, 1979; Loomis, 1982; Townsend, 1971; Van Der Heijden, Malhas, & Van Den Roovaart, 1984). For any given letter, one can average that letter's confusabilities with the remaining 25 letters of the alphabet to obtain a measure of that particular letter's confusability (LC). In addition, for any given string of letters, one can compute its overall confusability (Total Letter Confusability, or TLC) as the sum of the confusabilities of the individual letters in the string. High TLC letter strings will be thus strings of letters mostly composed of high-confusable letters, whereas low TLC letter strings will be strings of letters mostly composed of low-confusable letters. Moreover, the mean of the confusabilities of the letters in the string (Mean Letter Confusability) can be calculated.

It seems to us that LC could be an important factor to consider in experiments involving degraded presentation of letter strings. When stimulus quality is manipulated and stimuli are degraded (usually by reducing the contrast between the stimulus and the background) the letter comprising the string in input are difficult to identify and the visual similarity between letters might thus influence letter identification. The letter confusability could hence be a relevant factor to consider in researches analyzing the effects of stimulus degradation on speeded nonword reading in skilled readers, since degradation could have a stronger effect on nonword reading when LC is high than when it is low. In other words, it might be the case that more similar/confusable letters suffer more when degraded than less similar/confusable letters. If our hypothesis is plausible, it might be therefore important to match TLC across conditions in the kind of experiments involving the factorial manipulation of SQ that refute cascaded processing, a possibility which the authors did not consider in these studies.

The importance of letter similarity in visual word recognition is not totally new. The involvement of LC in reading emerged in fact from recent findings showing that this variable influences the performance of patients with pure alexia, also known as letter-by-letter (LBL) reading (e.g., Arguin & Bub, 2005; Arguin, Fiset, & Bub, 2002; Fiset, Arguin, Bub, Humphreys, & Riddoch, 2005; Fiset, Arguin, & McCabe, 2006). This deficit is associated with a damage affecting the left fusiform gyrus – a region in the temporo-occipital cortex – or the fibres conducting visual information to this region (e.g., Beversdorf, Ratcliffe, Rhodes, & Reeves, 1997; Binder & Mohr, 1992; Damasio & Damasio, 1983; Dejerine, 1892). The main behavioural feature of patients affecting from LBL dyslexia is very slow reading characterized by an abnormally large word length effect. In other words, LBL patients usually show a linear increase in the time required to recognize a word as a function of the number of letters it comprises (see, e.g., Patterson & Kay, 1982); several studies reported that, depending on the patient, the time required to read a word can increase from

500 ms to several seconds for each additional letter in the stimulus. On the contrary, unimpaired skilled readers read words of different length at a substantially invariant rate (Weeks, 1997). These data have been typically interpreted by suggesting that whereas skilled readers are able to recognize several letters simultaneously, LBL patients have lost this ability and instead decode words as a sequence of isolated letters, without any access to a spatially parallel process.

Recent findings clearly showed that the visual similarity among letters has a central role in LBL readers performance. In normal condition, as word length increases, so does the sum of the confusabilities of the constituent letters; an effect of word length may be thus due to the TLC (that is usually higher for longer letter strings) rather than to the number of letters in the stimulus. Consistent with this interpretation, Fiset et al. (2005) demonstrated, in fact, that the word length effect usually showed by LBL dyslexic patients disappears when TLC is balanced across word length, i.e. when the short and the long words presented in the task are matched in terms of TLC. These results have been interpreted by suggesting that LBL reading is due to a visual encoding impairment affecting letter recognition and that TLC affects reading performance by modulating the signal-to-noise ratio at the level of letter identification, a ratio that is abnormally low in LBL dyslexic readers. This finding also falsified the classical view explaining the pure alexia as a condition characterized by the absence of parallel processing, since the abolition of the word length effect under the appropriate condition provides evidence for residual parallel letter processing in these patients, even if this processing is highly susceptible of the negative impact of LC.

Furthermore, previous studies documented that, whereas LC has an effect for LBL dyslexic patients, this variable does not influence the behaviour of neurologically intact readers in standard viewing condition (e.g., Arguin et al., 2002; Fiset, Arguin, & Fiset, 2006). Nevertheless, we suggest that LC might have an effect on skilled readers performance when contrast is reduced. Partial support to this interpretation comes from findings proving that the word length effect showed by LBL dyslexic patients can be reproduced in skilled readers when the stimuli are degraded in the task (Fiset, Arguin, & Fiset, 2006; see also Fiset, Gosselin, Blais, & Arguin, 2006); the authors suggested that the visual impairment affecting letter recognition in LBL reading would be in fact simulated in normal readers by reducing the stimulus quality. As a consequence, since LC has a role in LBL readers performance and the LBL impairment can be simulated in normal readers by reducing stimulus quality, the effect of LC might become significant for skilled readers in degraded condition.

The goal of the present study has been to test this hypothesis. In particular, our experiment is directed to determine whether the letter confusability has a role for unimpaired subjects when reading nonwords. To this end we collected data from skilled Italian readers by presenting high-

TLC nonwords and low-TLC nonwords both in clear and degraded conditions. The hypothesis is that the effect of degradation would be larger for high-TLC nonwords than for low-TLC nonwords.

3.2 Method

Participants. Thirty students at the Università degli Studi di Padova who had Italian as their first language and normal or corrected-to-normal vision participated as volunteers.

Design. The experiment consisted of a 2x2 within-subjects design with Total Letter Confusability (TLC; low TLC vs. high TLC) and stimulus quality (SQ; clear vs. degraded conditions) as factors.

Materials. A total of 100 seven letter nonwords was selected as stimuli (these stimuli can be seen in the Appendix, section B). They were all pronounceable. Also, none of the nonwords had either orthographic or phonological neighbours. These stimuli were divided into two groups of 50 nonwords according with their TLC (high vs. low). Letter confusability was determined by averaging empirical letter-confusion matrices obtained in previous studies (Gilmore et al., 1979; Loomis, 1982; Townsend, 1971; Van Der Heijden et al., 1984)[16]; the TLC was calculated as the sum of the confusabilities of the letters comprising the string. Mean TLC values were 2.6 and 3.7 ($t(98) = 20.3$, $p < .001$) for the nonwords belonging to the low-TLC and high-TLC conditions respectively. The low TLC nonwords were divided into two groups of 25 nonwords balanced in terms of mean TLC values (2.6 vs. 2.6, $t<1$), *lowTLCa* and *lowTLCb*. Similarly, the high TLC nonwords were divided into two groups of 25 nonwords balanced for mean TLC values (3.7 vs. 3.7, $t<1$), *highTLCa* and *highTLCb*. *LowTLCa*, *lowTLCb*, *highTLCa*, and *highTLCb* were balanced in terms of initial phoneme. These four lists were created to counterbalance high and low TLC stimuli with SQ across participants: each participant saw 50 stimuli clear (25 low and 25 high in terms of TLC) and 50 stimuli degraded (25 low and 25 high in terms of TLC). The assignment of stimuli to the four conditions was counterbalanced across participants, i.e. if participant X saw *lowTLCa* and *highTLCa* stimuli clear and *lowTLCb* and *highTLCb* stimuli degraded, participant X+1 saw *lowTLCb* and *highTLCb* stimuli clear and *lowTLCa* and *highTLCa* stimuli degraded.

[16] From the averaged confusion matrix, the diagonal has been removed (see Arguin, et al., 2002) and the LC vector has been computed by summing up the letter-by-letter confusion values. The LC values ranged from .27 (letter L) to .71 (letter B), with an average of .48.

Apparatus. The experiment took place in a sound attenuated and dim lit room. Stimuli presentation and data recording were controlled by software developed in E-prime and running on a personal computer. The display was synchronized with the screen refresh cycle. Stimuli were presented centrally in upper-case letters in 18-point Courier New font on a black background (RGB values; 0, 0, 0). Clear stimuli were displayed in white (RGB values: 65, 65, 65); degraded stimuli were displayed in grey (RGB values: 8, 8, 7). Responses were collected via a microphone connected to a voice-key assembly. Response latency was timed from stimulus onset to voice key activation, which also terminated the display.

Procedure. Participants were tested individually and sat in front of the computer screen. They were instructed to read each letter string aloud as quickly as possible, and to minimize errors; they were informed that the stimulus would be a pronounceable nonword. Subjects were then presented with 12 practical trials. Each trial began with a 500 ms presentation of a fixation point at the centre of the computer screen followed by a 200 ms presentation of a blank. Immediately after the stimulus appeared and remained on the screen until a response was registered by the voice key or 3 sec elapsed. The inter-trial-interval was set to 2 sec. Stimuli were presented in a random order for each participant. The experimenter coded the pronunciation as correct if the pronunciation obey to the standard grapheme-phoneme rules, voice key failure or articulation error.

3.3 Results

Pronunciation errors (10.6%) and apparatus failures (4.98%) were removed prior to reaction times analysis. Correct reaction times were submitted to the Van Selst and Jolicoeur’s (1994) outlier removal procedure. Outliers (1.3%) were removed prior to RTs analysis. Mean RTs according to conditions and percentages of error are reported in Table 3.

	Stimulus Quality			
	Degraded		Clear	
TLC	RTs	%E	RTs	%E
High	1189	17	874	6
Low	1125	18	871	7
Diff.	64	-1	3	-1

Table 3. Mean reaction times (RTs) and percentages of error (%E) according to conditions.

In the ANOVA for the participants (F1) TLC and SQ were repeated factors. In the ANOVA for items (F2) TLC was an independent factors and SQ was a repeated factor.

RTs. Analysis showed a main effect of SQ, $F1(1, 33) = 92.8$, $MSE = 29575$, $p < .001$, $F2(1, 98) = 1001.4$, $MSE = 4242$, $p < .001$, and a main effect of TLC, $F1(1, 33) = 12.3$, $MSE = 3069$, $p < .001$, $F2(1, 98) = 4.4$, $MSE = 8585$, $p < .05$. However, the two effects were qualified by a significant interaction, $F1(1, 33) = 8.7$, $MSE = 3543.9$, $p < .01$, $F2(1, 98) = 6.8$, $MSE = 4242$, $p < .01$. Paired comparisons – by participants (t1) and by items (t2) – revealed that whereas the effect of TLC was significant when the stimuli were degraded, $t1(33) = 3.6$, $p < .001$, $t2(98) = 2.8$, $p < .01$, it was not significant when the stimuli were clear, $ts < 1$.

Accuracy. Whereas the main effect of SQ proved significant, $F1(1, 33) = 83.5$, $MSE = .006$, $p < .001$, $F2(1, 98) = 47.7$, $MSE = .016$, $p < .001$, neither the effect of TLC, $Fs <1$, nor the interaction, $F1(1, 33) = 2.1$, $MSE = .003$, $p > .15$, $F2(1, 98) = 1.3$, $MSE = .016$, $p > .2$, were significant.

3.4 Discussion

This study was directed to analyze the effect of letter confusability in nonword reading, a variable defined as the visual similarity between letters driven by shared features. Our experiment clearly showed that nonword reading is influenced by letter similarity in particular experimental conditions. In fact, the results obtained proved the letter confusability role in nonword reading when stimuli are presented in low contrast (i.e., degraded condition). Instead, letter confusability doesn’t affect nonword reading when stimuli are presented in standard viewing condition (i.e., clear condition).

More specifically, when the Total Letter Confusability (i.e., the sum of the confusabilities of the letters in the string) and SQ are jointly manipulated in a nonword reading task, a significant interaction between the two factors is obtained, with the effect of degradation being larger for the high-confusable that for the low-confusable strings of letters. As hypothesized, high-TLC nonwords are thus harmed by stimulus degradation more than low-TLC nonwords.

This result has at least two important implications for the research on visual word recognition and reading aloud.

First, the role of TLC has been largely documented in LBL dyslexia (e.g., Arguin & Bub, 2005; Arguin, et al., 2002; Fiset, et al., 2005). When the visual impairment affecting word recognition that characterizes this disorder is simulated in normal readers through stimulus degradation (see Fiset, Arguin & Fiset, 2006), then TLC has an effect on skilled readers performance. A variable affecting LBL dyslexics' behaviour has been thus shown to influence unimpaired skilled readers in particular experimental conditions.

Second, our findings have important implications for the researches using factorial manipulations to analyze information processing in visual word recognition. As said, recent studies involving the manipulation of SQ together with another factor in nonword reading tasks (e.g., Besner & Roberts, 2003; Reynolds & Besner, 2004) have provided strong evidence against cascaded processing. In particular, these studies suggested that information processing in the reading system might be thresholded at least in a particular level when reading nonwords. We argue that these data might be due to a confounding with TLC and that any threshold in the reading system would be instead needed. In fact, if the letter confusability plays a role when stimuli are degraded, then any study that involves a manipulation of SQ should take this factor into account. In particular, the effect of degradation is shown to be stronger for high-confusable letter strings than for low-confusable letter strings; this means that, if the stimuli in the experiments involving SQ are not controlled for TLC, the results obtained could be due to a counfounding with this variable. Consider, for example, the additive effects of SQ and letter string length that have been reported by Besner and Roberts (2003). The TLC is not controlled for in this study. This means that, as letter string length increases, so does the TLC, i.e. since TLC is calculated as the sum of the confusabilities of the letters in the string it is likely to increase as the number of letters in the string increases. Thus, part of the increased RTs for the degraded long nonwords compared to the degraded short nonwords obtained in this study could have been due to the increased TLC rather than to the increased length. If so, the additive effects that have been observed might be due to a counfounding with this variable: if short and long nonwords would be matched on TLC, then the true result could be an interaction, with the effect of SQ being larger for short than for long nonwords, as predicted by the DRC model. This hypothesis has been tested in the study reported in the next chapter.

3.4.1 Computational modelling

The basic assumption of computational modelling is that computational accounts are sensitive to the same variables that humans are sensitive to. In the present experiment we showed that letter confusability influences skilled readers performance in particular experimental conditions. Hence, since human readers are sensitive to letter confusability (at least for degraded stimuli), we must require the DRC model to be too.

The confusability between letters depends on letters' similarity that, in turn, depends on the letters' font and case[17]. This means that the LC for human readers might differ from the LC for the DRC model. As said, LC for human readers is empirically obtained. But what about the LC for computational models? Since letter similarity in human reading is driven by shared features, we argue that LC may be calculated in computational models as the proportion of visual features that two letters have in common, by considering the font and case that have been implemented in that specific model. Specifically to the DRC, it derives the first levels of processing from the IA model (McClelland & Rumelhart, 1981) that assumes the upper-case font produced by Rumelhart (1970) and Rumelhart and Siple (1974). This font is illustrated in Figure 5.

Figure 5. The features used to construct the letters and the letters in the font assumed by the simulation program on which the IA model and the DRC model are based (McClelland & Rumelhart, 1982, p. 383, Figure 4).

In the Rumelhart-Siple font there are 14 line segments which can be used to represent any letter. Different letters are represented by different subsets of these 14 lines. Hence, one can measure the confusability between any two letters in the font used to represent letters in the DRC model as the proportion of the 14 features used to code letters which the two letters share. Consider, for example, the letters E and F. These letters have 12 features in common; since the total number of features (present or absent) each letter has is 14, their confusability will be 12/14 = .857.

[17] Letter confusability values used in all the experiments reported in this thesis have been determined by averaging empirical letter-confusion matrices obtained in previous studies (Gilmore et al., 1979; Loomis, 1982; Townsend, 1971; Van Der Heijden et al., 1984). These experiments employed upper-case stimuli but the fonts used were not always the same; however, the correlation among the matrices is always quite high and this could be interpreted as a partial independence of confusion from font in these studies. Given the theoretical importance of the font when SQ is reduced, however, the same font (i.e., 18-point Courier New) and case (i.e., upper-case) have been used in all the experiments involving degradation presented in this thesis.

For any pair of letters in the DRC model's font, a measure of confusability of that pair can be obtained; hence, one can measure the average confusability for each DRC letter (DRC-LC), and hence calculate the DRC-TLC for any letter string.

Is the DRC model sensitive to the TLC? Unfortunately, the DRC model in its actual formulation is unable to simulate the effect of TLC obtained for skilled readers. In fact, when two sets of nonwords differing in terms of DRC-TLC are run through the computational version of the model under the degraded condition, any effect due to this variable is obtained. There are, however, two features of the DRC model that may play a role in determining this unsuccessful result.

Critical in this context is the feed-forward connection from the visual feature level to the letter level. There are two parameters in the DRC model controlling this connection, one that regulates the activation of the letters which have the visual features in input and another that determines the inhibition of the letters that have not those features. Since letter confusability is driven by shared features in the model, the values of these parameters are certainly important for the simulation of any effect involving letter confusability. Currently, the parameter that regulates the inhibition between visual feature and letter units is much higher (i.e., 30 times greater) than the parameter that regulates their excitation[18]. This means that just one mismatching feature will completely block activation of similar letters. As a consequence, any effect due to letter similarity won't be simulated by the DRC model implementing these default values.

The other important parameter for the simulation of letter confusability effects is the Letter-Lateral-Inhibition. This parameter determines the inhibition that each letter at the letter level sends to the competitive letters. In the model, when a letter receives activation, it should inhibit all the other letters and, through this mechanism of lateral inhibition, the system would gradually converge on a single unit that corresponds to the target letter. Currently, the value of this parameter is zero: this means that different letters cannot inhibit each other in the model and multiple letters activation will therefore only interfere. To simulate any effect due to letter confusability, lateral inhibition between different letters is instead required.

Clearly, the current values of the parameters described above are inadequate and do not allow the present computational version of the DRC model to reproduce any effect due to letter similarity. However, one does not have to adhere to these values. It might seem indeed peculiar that the inhibition between feature and letter units is 30 times greater than the excitation between these two levels; more likely, these values should be the same. Moreover, letter-lateral inhibition is assumed in the model without be actually implemented, since the value of the parameter regulating

[18] Precisely, the default values of these parameters in the DRC model are .15 for the feature-to-letter inhibition and .005 for the feature-to-letter excitation.

this inhibition is zero: hence, a change of this value is clearly justified (and perhaps needed) from a theoretical point of view. Moreover, the values that the DRC model actually implements for these parameters have been inherited from its progenitor, the IA model (see Table 1, p. 387 of McClelland & Rumelhart, 1981); to the best of our knowledge, however, this precise setting is not required by any empirical data.

To conclude, we argue that our empirical findings claim a change of the values of some parameters assumed in the DRC model, i.e. the parameters regulating the connections between the feature and the letter units and the parameter regulating the lateral inhibition within the letter level[19]. According to the principle of nested modelling this is a plausible way to proceed, as long as the model so modified will be still able to reproduce the pattern of results that its actual computational version does simulate.

Importantly, our results require not only the DRC model, but more generally every computational model of reading to simulate the effects due to letter similarity when stimuli are degraded. Clearly, this is a challenge that should be taken into account by future works on computational modelling of the reading process.

3.4.2 Conclusion

A variable playing a role for letter-by-letter dyslexic patients – the Total Letter Confusability – has been shown to influence skilled nonword reading when stimuli are degraded. This finding has important implications for researches on visual word recognition: since high-TLC stimuli are harmed by stimulus degradation more than low-TLC stimuli, then any study that involves a manipulation of SQ should take the TLC into account.

[19] I will hark back on this issue in the final chapter of this thesis.

4 STIMULUS QUALITY AND LETTER STRING LENGTH IN NONWORD READING

In this chapter the joint effects of stimulus quality and letter string length in nonword reading will be analyzed. Besner and Roberts (2003) reported that, whereas these factors have additive effects on readers' latencies in a nonword reading task, they interact in simulations of the DRC model. The authors suggested that the DRC model would only be able to capture the additivity of length and degradation by a radical change to the model, namely thresholding the output of the letter level.

We argue that the results reported by Besner and Roberts (2003) may be instead due to a confounding involving a variable representing letter similarity – the Total Letter Confusability (TLC). Since TLC plays a role in degraded nonword reading, then any study involving a manipulation of SQ should match TLC across conditions. Following this hypothesis, we will show that SQ and letter string length interact in the reading task when short and long nonwords are matched for TLC. Implications for models of visual word recognition will be discussed.

4.1 Introduction

Skilled readers' latencies to nonwords increase monotonically as the number of letters increases (e.g., Weekes, 1997), thus suggesting that print is translated into sound serially along the string of letters when reading nonwords. This assumption is central in the DRC model of reading. According to the DRC model, in fact, nonwords are read through the non-lexical route, a serial procedure that, using language-specific correspondence rules, converts each grapheme into the corresponding phoneme from left to right. As the number of letters in a nonword increases, the time required to the non-lexical route to reach the rightmost letter and activate the last phoneme increases as well. The effect of letter string length in nonword reading is indeed correctly simulated by the DRC model (e.g., see simulations reported in Chapter 2).

The effect of letter string length when jointly manipulated with stimulus quality has been recently analyzed (Besner & Roberts, 2003). The goal of this study was to determine whether the effect due to letter string length in nonword reading changes in function of the stimuli being presented in a clear condition or in reduced contrast. Critically, additive effects of the two variables have been reported on reading latencies, i.e. the effect of letter string length has the same amplitude regardless of the levels of SQ.

Moreover, Besner and Roberts (2003) attempted to simulate these effects with the DRC model. Critically, SQ and letter string length interacted in this simulation[20], with the effect of letter length being larger for clear than for degraded nonwords. In other words, slowing the rate of processing affected short nonwords more than long nonwords in the DRC model's performance. This seemingly counterintuitive result is explain in a straightforward way by the model. As said, the DRC model engages serial processing for nonword reading. Thus, when the rate of processing is slowed down by reducing SQ, the first phoneme of the letter string is delayed. However, the delay associated with the start of activation of each additional phoneme decreases as the number of letter increases. In fact, activation is continuously accumulated during phonemic processing and, since reading longer nonwords requires more time, activation grows more for longer letter strings. Given that pronunciation does not start until all phonemes reach threshold, the delay produced by stimulus degradation is reduced for longer letter strings.

Critically, the pattern of results reported by Besner and Roberts (2003) showed a qualitative difference between the behaviour of human skilled readers and the simulations of the DRC model. According to the authors these data would call for a modification to the way the DRC model processes along the non-lexical route. In particular, the authors suggested that the interaction between SQ and nonword length can be eliminated by thresholding the output of a level somewhere in the model but reflecting early processing, i.e. either the visual feature level or the letter level. As the authors correctly observed, if the visual feature level would be thresholded, then the manipulation of processing rate will not affect anything beyond the feature level. As a consequence, several well-established two-way interactions in visual word recognition would not be explained by such an hypothesis. For example, a threshold at this level would be inconsistent with the interaction of SQ and repetition of words (i.e., words presented for the first time are more affected by degradation than words presented for the second time) that has been shown both in lexical decision (Besner & Swan, 1982) and in reading (Blais & Besner, 2007); in a similar way, such a threshold would not explain the interaction of SQ and semantic priming (i.e., the effect of semantic priming is larger when the word is degraded compared with when it is clear) also obtained both in lexical decision and reading tasks (e.g., Besner & Smith, 1992; Borowsky & Besner, 1993; Ferguson, Robidoux, & Besner, 2009; Meyer, et al., 1975; Stolz & Neely, 1995).

[20] As said, stimulus quality is usually simulated in computational models by modifying the connection weights so to reduce the processing rate in early levels. Specifically, Besner and Roberts implemented degradation in the DRC model by reducing the connections between the feature and the letter units by 40%.

A different possibility is to threshold the output of the letter level. In fact, the interaction between SQ and letter string length in nonword reading reflects the effect of cascaded processing at the letter level, followed by serial processing when translating letters into phonemes. If the letter level is thresholded, then the effect of reduced stimulus contrast would not affect the model beyond the letter level and, given that the nonword length effect arises from the subsequent serial assignment of phoneme to grapheme, the joint effects of SQ and letter string length would be additive on reading latencies. Computational simulations showed indeed this is the case (see Besner et al., 2003). The authors also proposed that thresholding the letter level would resolve other problems as well: such a modification would in particular explain the additive effects on reading latencies of SQ and orthographic neighbourhood (N) size (Reynolds & Besner, 2004; but see Blais & Besner, 2007) and of SQ and word frequency (O'Malley & Besner, 2008)[21].

Moreover, the authors suggested that the letter level would be thresholded only before it activates the non-lexical route and it instead cascades to the lexical route. In other words, whereas the grapheme-phoneme conversion process would be activated by the output of the letter level in a thresholded fashion, the lexical route would be activated by cascaded letter level processing (see also Blais & Besner, 2007). Given the interactive-activation between the orthographic lexicon and the letter level assumed in the model remains intact, this account is consistent with the interactions observed between SQ and word repetition (e.g., Blais & Besner, 2007), between SQ and semantic priming (e.g., Ferguson et al., 2009) and between SQ and word frequency when the reading task comprises only words (O'Malley & Besner, 2008; Yap & Balota, 2007).

Clearly, adding a threshold at the letter level means to accept that information processing in the reading system is – at least at one level of processing – discrete and serially organized (Sternberg, 1969; see also Sternberg, 1998). Accepting this proposal therefore requires a radical change of the DRC model, e.g. the C in DRC should be abandoned since C stands for cascaded and not thresholded. More generally, the idea itself that information processing in the reading system is cascaded would be refuted.

We argue that a different interpretation of the Besner and Roberts' (2003) results can be proposed by considering the Total Letter Confusability (or TLC). The letter confusability is, as demonstrated in the previous chapter, a measure of letter similarity which effects become significant in degraded reading. In particular, it has been shown that the more confusable a letter is to the other letters of the alphabet, the stronger the effect of degradation will be on that letter's identification. As a consequence, any study involving degraded nonword reading should take this measure into account.

---

[21] These issues will be further discussed, respectively, in Chapter 5 and in Chapter 6.

Critically, Besner and Roberts (2003) didn't match their stimuli for TLC. Hence, since TLC usually increases as letter length increases, our assumption is that the long nonwords used in their experiment had higher TLC values than their short nonwords. Moreover, an analysis on the stimuli used by the authors confirms this interpretation. Their experiment used lowercase letters and, to the best of our knowledge, there is only one published confusion matrix for lowercase letters, that of Courrieu, Farioli, and Grainger (2004). Our analysis of the material used in the Besner and Roberts' (2003) experiment indicated that their long nonwords had a much higher mean TLC than their short nonwords (551.7 vs. 376.8; $t(62) = 13.9$, $p < .001$). Therefore part of the increased RTs for the degraded long nonwords compared to the degraded short nonwords could have been due to the increased TLC rather than to the increased length. If so, then matching the long and short nonwords on TLC would reduce the difference in RTs between the degraded long nonwords and the degraded short nonwords; that would reduce the slope of the length effect in the degraded condition, so that the effects of degradation and length would no longer be additive. Instead, the length effect would be smaller for degraded than for clear stimuli: which is the effect to be expected from the DRC model.

The experiment reported below was directed to test this possibility by matching short and long nonwords on TLC, thus eliminating the confounding due to this variable. A condition in which TLC was not controlled has been also introduced to replicate the Besner and Roberts' (2003) study.

4.2 Method

Participants. Thirty students at the Università degli Studi di Trento who had Italian as their first language and normal or corrected-to-normal vision participated as volunteers.

Design. The experiment consisted of a 2x2x2 within-subjects design with Total Letter Confusability (TLC; balanced vs. unbalanced conditions), length (short vs. long items), and stimulus quality (SQ; clear vs. degraded conditions) as factors.

Materials. A total of 120 pronounceable nonwords was selected (these stimuli can be seen in the Appendix, section C). The nonwords had neither orthographic nor phonological neighbours[22].

[22] Note that nonwords with an orthographic neighborhood size of 2 were used in the Besner and Roberts' (2003) study; the authors, however, interpreted the effects obtained as caused by a purely nonlexical processing: hence the use of nonwords without neighbours in our experiment.

The nonwords were divided into two groups of 60 stimuli belonging to two levels of Total Letter Confusability (balanced vs. unbalanced conditions). Within each group there were 30 stimuli at each of two levels of length (short vs. long) matched for their initial phoneme. Items that were 5 letters in length represented the short length condition; items that were 7 letters long represented the long length condition. Short and long nonwords were balanced in terms of TLC (2.69 vs. 2.69, $t<1$) in the balanced condition and significantly differed for this factor (2.17 vs. 3.43, $t(58) = 19.2$, $p<.001$) in the unbalanced condition. Letter confusability has been determined from previous empirical letter-confusion matrices (Gilmore et al., 1979; Loomis, 1982; Townsend, 1971; Van Der Heijden et al., 1984). The TLC was calculated as the sum of the confusabilities of all letters in an item. Finally, at each of two levels of length, half the items were presented in the clear condition and the other half in the degraded condition. Items were counterbalanced across levels of stimulus quality in such a way that half the subjects saw an individual item under the clear (degraded) condition and the remaining subjects saw that item under the degraded (clear) condition.

Apparatus and Procedure. The same apparatus and procedure of the experiment reported in Chapter 3 have been used.

4.3 Results

Pronunciation errors (10.5%) and apparatus failures (12.9%) were removed prior to reaction times analysis. Correct reaction times were submitted to the Van Selst and Jolicoeur's (1994) outlier removal procedure. Outliers (2.7%) were removed prior to RTs analysis. Mean RTs according to conditions and percentages of error are reported in Table 4.

	Total Letter Confusability							
	Balanced				Unbalanced			
	Stimulus Quality				**Stimulus Quality**			
	Clear		Degraded		Clear		Degraded	
Length	RTs	%E	RTs	%E	RTs	%E	RTs	%E
Long	806	13	859	19	789	15	877	17
Short	687	6	768	9	681	8	747	10
Diff.	119	7	91	10	108	7	130	7

Table 4. Mean reaction times (RTs) and percentages of error (%E) according to conditions.

In the ANOVA for the participants (F1) TLC, SQ, and Length were repeated factors. In the ANOVA for items (F2) TLC and Length were independent factors, SQ was a repeated factor.

RTs. Analysis showed a main effect of Length, $F1(1, 29) = 154.9$, $MSE = 4839$, $p < .001$, $F2(1, 232) = 166.9$, $MSE = 4490$, $p < .001$, and a main effect of SQ, $F1(1, 29) = 70.1$, $MSE = 4452$, $p < .001$, $F2(1, 232) = 74$, $MSE = 4490$, $p < .001$. As in the Besner and Roberts' (2003) study, SQ and Length do not interact, $Fs < 1$. However, the three ways interaction among TLC, SQ, and Length proved significant, $F1(1, 29) = 10.4$, $MSE = 844$, $p < .005$, $F2(1, 232) = 6.3$, $MSE = 4490$, $p < .05$. When TLC is controlled for across length, the length effect for degraded stimuli is smaller than the length effect for clear stimuli (91 vs. 119 ms, respectively); whereas when TLC is left uncontrolled, the length effect for degraded stimuli is bigger than the length effect for clear stimuli (130 vs. 108 ms, respectively).

Errors. Only the main effects of SQ, $F1(1, 29) = 6.2$, $MSE=.012$, $p < .05$, $F2(1, 232) = 5.4$, $MSE=.014$, $p < .05$, and Length, $F1(1, 29) = 31.8$, $MSE=.0102$, $p < .001$, $F2(1, 232) = 23.3$, $MSE=.014$, $p < .001$, proved significant.

4.4 Discussion

This study showed three principal results.

First, long nonwords are read slower than short nonwords. This effect is well-established in reading researches (e.g., Weekes, 1997). In the DRC model this effect is explained because nonwords are assembled letter-by-letter (from left to right) by the non-lexical route.

Second, we found an effect of degradation, due to the fact that clear stimuli are read aloud faster than degraded stimuli. This result is due to degradation slowing the rate of processing in the reading system. In the DRC, a reduction in SQ is implemented by reducing the weights of the connections between the feature and the letter units: the effect of stimulus quality is correctly simulate by the DRC model (see, e.g., Besner & Roberts, 2003).

Third and most important, the three ways interaction between SQ, length and TLC proved significant. In particular, we showed that when TLC is controlled for across length, the length effect for degraded stimuli is smaller than the length effect for clear stimuli. Instead, when TLC is left uncontrolled, the length effect for degraded stimuli is larger than the length effect for clear stimuli.

We therefore conclude that the additivity of degradation and nonword length reported by Besner and Roberts (2003) occurred because of a confounding between TLC and length, and that when this confounding is removed the two factors interact, with the length effect being smaller for degraded stimuli than for clear stimuli, a result which, as Besner and Roberts (2003) showed, is also simulated by the DRC model.

4.4.1 Computational modelling

An issue concerning computational modelling remains however to be resolved. Even it were true that the additivity of degradation and length observed in human reading by Besner and Roberts (2003) occurred because of a confounding between TLC and length, the DRC model ought still to have been able to simulate it, because it is supposed to be sensitive to the same variables that human readers are sensitive to. So the requirement that the DRC model be able to produce an additivity of degradation and length with the Besner and Roberts' stimuli has not been avoided.

As said in the previous chapter, a first problem in simulating letter confusability effects might regards how TLC is calculated. In fact, the LC for human readers with the font and case used by Besner and Roberts (2003) may differ from LC for the DRC model with the font and case used by this model. Hence, if the DRC model does not produce additivity of degradation and length when the Besner and Roberts' stimuli are used, this might be because, for the DRC's font, there is no difference in TLC between the short and the long nonwords used by Besner and Roberts and hence no confounding of length with TLC. However, an analysis on the stimuli used by Besner and Roberts (2003) turns out that the DRC-TLC is significantly higher for the long items than for the short items (3.22 vs. 2.23; $t(62) = 13.5$, $p < .001$) used in this study. Hence, the confounding of length with TLC is significant for the DRC model as for human readers and the additive effects reported by Besner and Roberts (2003) should be correctly simulated by the model.

In the previous chapter we identified two parameters of the DRC model that are certainly involved in any effect depending on letter confusability. As said, the actual values of these parameters are not adequate and, as a consequence, the DRC model cannot simulate the effect due to TLC we showed in human reading; moreover, the DRC model in its actual setting won't reproduce any result that depend on this variable. The task will be therefore to find the right manipulation of these parameters that shows additivity of SQ and letter string length when TLC is confounded with letter length but also an interaction between the two factors with smaller length effect for degraded than for clear stimuli when TLC is matched across length. Only if both these results are obtained the DRC model can be considered successful.

This issue will be further discussed in the final chapter of this thesis, where a few attempts of simulation in this direction will be successfully presented.

4.4.2 Letter's position encoding: evaluating a theoretical account

Besides the issues here discussed, the researchers analyzing the joint effects of SQ and letter string length might be useful to evaluate a novel theoretical framework proposed to explain how the position of a letter within a string is encoded, the Sequential Encoding Regulated Inputs to Oscillations within Letter units (SERIOL) model (Whitney, 2001).

Computational models of visual word recognition need specific assumptions about how letter position is encoded. In the DRC model, for example, different sets of letter units exist for each string position. In other words there are, in the model, separate units that represent the letter *A* in the first position, the letter *A* in the second position and so on (see also Coltheart, Curtis, Atkins, & Haller, 1993; McClelland & Rumelhart, 1981; Whitney, Berndt, & Reggia, 1996). Thus, for example, the word *ART* is encoded in the model by activating *A* in the first subset, *R* in the second subset and *T* in the third subset. This organization certainly demands a high degree of redundancy since a representation of each letter in each possible position is required[23].

An alternative neurobiologically plausible theoretical framework assuming serial processing – the SERIOL model – has been recently proposed to account for letter position encoding.

Briefly, the first level assumed in the model, the retinal level, correspond to the earliest level of visual processing. Units at this level correspond to pixel and are topographically organized with respect to external stimuli. The representation of the string in input is split across the hemispheres, so that the left visual field (LVF) is processed by the right hemisphere (RH), whereas the right visual field (RVF) is processed by the left hemisphere (LH). In the physical retina visual acuity decreases with increasing angle from the fixation point (due to the decreasing concentration of cones); in a similar way, the activation of the retinal units in the model – the acuity gradient – decreases as distance from fixation increases. The acuity gradient is thus symmetric across fixation, with decreasing activation from the fovea to periphery, i.e. the acuity gradient increases form the first letter to the fixation (i.e., in the LVF/RH) whereas it decreases from fixation to the last letter (i.e., in the RVF/LH).

[23] Alterative accounts are also available. For example, to reduce redundancy it has been proposed that each unit may represent not only the letter but also its position in the string, e.g. *ART* would be represented by *A-1*, *R-2*, *T-3*. However, it is unclear how this tagging could be realized in physiological terms. Or else, it has been suggested that the basic units may be groups of ordered letters such as trigram rather than single letter (Seidenberg & McClelland, 1989; see also Mozer, 1987), e.g. the word *ART* would be coded as *_AR*, *ART*, *RT_*, (where _ represents a word boundary).

At the subsequent feature level, the acuity gradient is converted into an activation pattern – the spatial gradient – that decreases across the letter string from left to right. Therefore, the slope of the LVF/RH acuity gradient is in the opposite direction as required for spatial gradient, while the slope of the RLF/LH acuity gradient is in the same direction. Thus, in the LVF/RH, the acuity gradient slope must be inverted as it activates letters' features; in contrast, the acuity gradient slope can be maintained as features are activated in the RVF/LH. As a result, processing at the feature level differs across hemisphere. This hemisphere-specific processing is assumed to be learned during reading acquisition, probably in response to attentional mechanisms. In particular, excitation from the retinal level to the feature level is assumed to be strong in the LVF/RH. This allows the first letter's features to reach a high level of activation in the LVF/RH even if it's far from fixation[24]. Also, strong directional lateral inhibition connections within the feature level are assumed in the LVF/RH such that each feature unit inhibits units to its right so strongly to invert the slope of the gradient. On the contrary, excitatory and lateral inhibition connections are weaker in the RVF/LH because the slope of the acuity gradient is already in the correct direction. In addition, LVF/RH features inhibit the RVF/LH features, bringing the activation of the latter lower than activation of the former. The two parts of spatial gradient are finally combined through inter-hemispheric callosal transfer creating an activation gradient decreasing from the first letter to the last letter.

At the next level, the letter level, the spatial gradient induces a temporal firing pattern across letter units. Specifically, due to the location gradient, the letter node representing the letter in the first position receives the highest level of excitatory input, the second receives the next highest amount, and so on. Letter nodes receiving the highest levels of input will fire first because reach threshold before the others; also, lateral inhibition ensures that only one letter node fires at a time. Hence, in the SERIOL model, letter's position is represented by the precise timing of firing of a letter node relative to the other letter nodes.

Finally, a bigram level and a word level are also assumed in the model. However, their description is superfluous for the present purposes.

The length effect is simply accounted for in the SERIOL model, given that letters are serially decoded. In particular, the length effect would be due to the fact that longer strings are presented farer on the left side of fixation than shorter strings; for letters on the left side of fixation the natural acuity gradient (i.e., visibility degrading from the fovea to periphery) must be reversed by strong excitation and left-to-right later inhibition assumed in the LVF/RH. Specifically, through

[24] This assumption is consistent with empirical data showing that perceptibility of initial letters does not decrease as distance from fixation in the LVF/RH; on the contrary, it does decrease in the RLV/LH, where activation from the retinal level is lower (Estes, Allemeyer, & Reder, 1976).

the directional later inhibition, each feature inhibits the units to its right so strongly to invert the slope of acuity gradient. Critically, inhibitory input increases as letter-position increases, because more and more features will send inhibition from the left. As consequence of this mechanism, the LVF/RH spatial gradient becomes more and more non-linear as the number of letters on the left size of fixation increases; hence, activation will be reduced across letter-position delaying letter fairing for longer string and increasing the amount of time required for the network to reach criterion, hence producing a length effect.

As consequence of the logic described above, the length effect depends on spatial gradient formation in the LVF/RH. Consistently with this prediction, many studies showed that the length effect is obtained for stimuli parafoveally presented in the LVF but it disappears when the stimuli are presented in the RVF (e.g., Bouma, 1973; Ellis, Young, & Anderson, 1988; Melville, 1957; Young & Ellis, 1985). In fact, the perceptual span in the LVF is four letters (Rayner, 1975); thus, in LVF parafoveal presentation, the letters of a long string are not maximal activated by their features because bottom-up input is lower than for central fixation. In addition, strong lateral inhibition from the first letter has a large effect on the second and third letter, because of their low level of bottom-up input. Due the lower activation of the initial letters and the increasing activation levels of the final letters (due to acuity increasing near fixation), lateral inhibition would thus fail to create a smoothly decreasing spatial gradient. On the contrary, in RVF/LH parafoveal presentation, the spatial gradient remains smoothly decreasing because it is based on the acuity gradient. Moreover, Whitney and Lavidor (2004) showed that the length effect obtained in the LVF was annulled by increasing 2nd and 3rd letter contrast (in 4 and 6 letter long words) whereas the same manipulation made the effect appearing in the RVF. The idea is that increasing 2nd and 3rd letter contrast would increase bottom-up input of these letters, hence creating a smooth gradient in the LVF/RH that, facilitating longer strings, would cancel the length effect. Conversely, the application of the same pattern in the RVF/LH should create a length effect due to disruption of a previously smooth spatial gradient.

We argue that the SERIOL model also makes clear predictions regarding the joint effects of SQ and letter string length. As said, the length effect is due to an attentional pattern that reverses the natural visibility gradient by increasing the bottom up input of the left part of the string. Since SQ influences the amount of bottom up input, stimulus length and degradation exert their effects – at least in part – at the same level in the SERIOL model. Hence, the model predicts that SQ and letter string length interact, with the length effect being larger for degraded than for clear stimuli. In particular, reducing stimulus contrast will cause the production of a non-linear spatial gradient, which will slow long letter strings more than short letter strings. In degraded presentation, in fact, the letters in the LVF won’t be maximal activated by their features given the reduced bottom up input. As a consequence, lateral inhibition from the first letter will have a large effect on the subsequent letters at the left of fixation (because of their low level of activation), thus creating a non-linear spatial gradient. Since inhibitory input increases as letter-position increases (because more and more features send inhibition from the left) the spatial gradient will become more and more non-linear as the number of letters in the string increases. Hence, the effect of degradation will be stronger for long letter strings compared to short letter strings.

Critically, the results expected by the SERIOL model are opposite to the predictions of the DRC model. In fact, the effect of letter string length is predicted to be larger for degraded than for clear stimuli by the SERIOL model; instead, the length effect is expected to be smaller for degraded than for clear stimuli according to the DRC model. Crucially, the results obtained in our experiment showed an interaction consistent with what expected by the DRC model, thus falsifying the prediction of the SERIOL model. Our finding may therefore be relevant also to distinguish between these different accounts.

To conclude, the results we obtained are inconsistent with a novel theoretical framework of how the position of a letter within a string is encoded, the SERIOL model. This theory models visual word recognition from the retinal level to the lexicon and has been developed in order to be consistent not only with psychological studies but also with current theories of neural computation and physiology. Despite our results falsified the SERIOL model, we believe that the development of neurobiologically plausible accounts is certainly interesting and might be promising for the future development of researches in visual word recognition.

4.4.3 Conclusion

Besner and Roberts (2003) carried out an experiment on speeded nonword reading in which they varied two factors: nonword length, and whether stimulus presentation was clear or degraded. In their data, these two factors had additive effects on nonword reading latencies. They reported that the DRC computational model of reading did not correctly simulate this additivity. When the same nonwords were presented to the DRC model for reading, and degradation was simulated by reducing the strength of the excitatory and inhibitory connections from visual features to letters, length and degradation interacted: the DRC model’s latencies showed a smaller effect of length when the nonwords were degraded than when they were clear. Besner and Roberts (2003) suggested that the DRC model would only be able to capture the additivity of length and degradation by a radical change to the model, i.e. thresholding the output of the letter level to the non-lexical route. We demonstrated instead that this additivity is due to a counfounding with the Total Letter Confusability, a variable that we showed to be involved in reading when stimuli are degraded (see Chapter 3). In fact, when TLC is controlled for across letter string length, the length effect for degraded stimuli is smaller than the length effect for clear stimuli, as predicted by the DRC model. As a consequence, the results obtained by Besner and Roberts (2003) do not require a threshold in the reading system; instead, the results expected by cascaded processing are obtained under the appropriate experimental condition.

5 STIMULUS QUALITY AND ORTHOGRAPHIC NEIGHBOURHOOD SIZE IN NONWORD READING

In this chapter, the neighbourhood size (N; Coltheart et al., 1977) effect will be analyzed when nonwords are presented both in a clear condition and in a degraded condition in a reading aloud task. This type of manipulation is interesting in visual word recognition and reading aloud researches at least for two reasons. Fist, the manipulation of stimulus quality in combination with other factors has been used within discrete stage accounts to delineate the processing sequence in reading; in particular, it has been suggested that the joint manipulation of SQ and N might have important implications in determining the locus of the N effect in reading. Second, previous studies (e.g., Reynolds & Besner, 2004) analyzing the joint effects of SQ and N both in skilled readers and in the DRC computational model reported that, whereas SQ and N exert additive effects on skilled readers latencies, the two factors interact in DRC model simulations, with the effect of SQ being smaller for high-N nonwords than for low-N nonwords; the results observed for human readers appear therefore to be inconsistent with cascaded processing assumed in the model and a threshold at the letter level has been proposed as a solution.

The aim of the present study is, from one hand, to test the hypothesis of a threshold in the reading system and, from the other, to provide further evidence regarding the locus of the N effect in nonword reading.

5.1 Introduction

An important question in reading researches is whether and how lexical knowledge affects nonword reading. One approach to answering this question involves examining the orthographic neighbourhood density (N; Coltheart et al., 1977), an effect arising from activation within the lexical route (e.g., Andrews, 1997; Coltheart et al., 1977, 2001; Reynolds & Besner, 2004). Regardless of many studies provided strong evidence in favour of the N effect in nonword reading (e.g., Andrews, 1989; McCann & Besner, 1987; Peereman & Content, 1995; Sears, Hino, & Lupker, 1995), the locus at which this effect arises in the reading system is still not clearly defined.

One hypothesis is that the N effect arises early in processing, through the interactive activation between the orthographic lexicon and the letter units (Andrews, 1989; see also Sears et al., 1995). According to this interpretation, the lexical entries corresponding to the neighbours of the stimulus in input would be activated in the lexicon and in turn would facilitate target letters' identification through the feedback activation from the orthographic lexicon to the letter level.

A second account is in favour of a late locus: the N effect would be due to the feed-forward connections from the orthographic lexicon to the phonological lexicon (Peereman & Content, 1995). The N effect would thus arise in reading because the orthographic lexical knowledge feeds forward to later phonological processes, thus facilitating the computation of phonology.

The issue concerning the locus of the N effect has been investigated within the DRC model. Despite Coltheart et al. (2001) initially suggested that the effect of N occurs late in the reading system, Reynolds and Besner (2002) demonstrated that there are both an early locus and a late locus of the N effect in the DRC model when reading nonwords. The authors performed several simulations through the DRC model proving that the lexical route can influence nonword reading both through the interactive activation between the letter units and the orthographic lexicon and through the feed-forward connections from the orthographic lexicon and the phoneme system. In fact, the DRC model still produces the effect of N in nonword reading when lesioned eliminating the connections into and out of the phonological lexicon, i.e. the only lexical contribution to performance when reading nonwords arises from the feedback between the orthographic lexicon and the letter units (i.e., early locus). In addition, the DRC model still produces a robust N effect when the feedback connections to the letter units are lesioned so that the only lexical contribution to nonword reading arises from the feed-forward connections to the phoneme system (i.e., late locus).

Nevertheless, the previous investigations didn't clarify whether skilled readers are affected by N at an early level or at a late level when reading nonwords aloud.

Reynolds and Besner (2004) suggested to analyze this issue by jointly manipulating N and stimulus quality in the task. The authors assumed in fact that the effect of a reduction in SQ occurs early in processing and that factors that interact with SQ would arise somewhat early in the reading system, whereas factors that are additive with SQ would have their effect later[25]. Hence, according to this interpretation, an interaction between SQ and N would support an early locus of the N effect, whereas additive effects of the two factors would be consistent with a late locus: since the empirical

[25] According to the authors this would be the simplest way to understand the following findings: 1. SQ and word frequency have additive effects on RTs (e.g., Balota & Abrams, 1995; Borowsky & Besner, 1993) 2. SQ interacts with semantic and repetition priming (e.g., Besner & Smith, 1992; Borowsky & Besner, 1993) 3. Semantic and repetition priming interact with word frequency (e.g., Becker, 1979; Visser & Besner, 2001). However, since interactions between SQ and word frequency have also been documented in reading (e.g., O'Malley & Besner, 2008; Yap & Balota, 2007), we argue that one should be cautious to come to a similar conclusion.

investigation reported that SQ and N exert additive effects on nonword reading latencies, the authors concluded a late account of the N effect for skilled readers. However, this interpretation clearly requires that at least some processes in reading are discrete and serially organized.

To date, the fact that orthographic neighbourhood has its effect late in processing is suggested also by multiple experiments reported by Reynolds and Besner (2006) using the psychological refractory period (PRP) paradigm. In the PRP paradigm (see Pashler, 1994) subjects perform two speeded tasks (Task 1 and Task 2) presented in close succession and are typically instructed to answer to Task 1 first. The interval between the two tasks (stimulus onset asynchrony, or SOA) is manipulated and the typical finding is the so called PRP effect, i.e. as SOA decreases, the time to respond to Task 2 increases. Many theorists (e.g., Pashler, 1984; Welford, 1952) ascribe this delay to both tasks using the same limed-capacity attention mechanism, or central attention (see Johnston, McCann, & Remington, 1995): if Task 2 requires the same process of Task 1, it is functionally postponed until that process becomes available[26]. According to this logic, the PRP effect would have straightforward consequence in determining whether processes involved in Task 2 occur before, during, or after the bottleneck. When Task 1 and Task 2 overlap temporally and subjects are instructed to respond to Task 1 before Task 2, Task 2 would be postponed; if a factor manipulated in Task 2 occurs prior the processing bottleneck, then the effect of this factor should be partially absorbed into the slack created by Task 2 processing waiting for central attention to become available. Hence, the effect of the factor manipulated in Task 2 will be underadditive with decreasing SOA, i.e. the effect of the factor manipulated in Task 2 will be smaller or absent at shorter SOA (but see Besner, Reynolds, & O'Malley, 2009). On the contrary, if a factor manipulated in Task 2 affects a process that occurs either during or after the bottleneck, it will have additive effects with SOA, i.e. the size of the effect produced by the factor manipulated in Task 2 will not be modulated by the length of the SOA. The experiments conducted by Reynolds and Besner (2006) employed a tone identification task (Task 1) followed by a reading aloud task (Task 2); an underadditive interaction between long-term repetition priming[27] and SOA has been observed when reading aloud in Task 2, thus suggesting that representations in the orthographic lexicon are activated prior the bottleneck; in contrast, additive effects of N and SOA were obtained, suggesting that N has its effect at or after the bottleneck. According to the authors, these findings would imply that N has its effect late in processing, i.e. after the activation of entries in the orthographic lexicon.

---

[26] Theoretical variants also exist: these accounts generally assume that some processes share capacity between Task 1 and 2 rather than an all-or-none bottleneck (e.g., Navon & Miller, 2002).

[27] Long-term repetition priming refers to faster performance for repeated items relative to novel items over lags greater than 100 intervening items. Since it is observed for words but not for nonwords, it is not affected by change in case and it interacts with word frequency (i.e., the effect is larger for low-frequency words than for high-frequency words), the long-term repetition priming effect is considered to affect lexical encoding.

The conclusion that N has a late effect in the human performance is clearly conflicting with the account proposed within the DRC framework. Further inconsistencies emerged from the analysis of the joint effects of SQ and N in the DRC model (see Reynolds & Besner, 2004). In fact, when SQ and N are jointly manipulated in the DRC model simulations, the two factors interact, with the effect of stimulus degradation significantly larger for low-N stimuli than for high-N stimuli. This interaction has been interpreted by suggesting that N has an early effect in the DRC model; however, given the assumption that an early effect also affects processing downstream in a cascaded model as the DRC, this result is also consistent with a later effect.

In this study we focus on the joint manipulation of SQ and N in a reading aloud task. This issue is central for the rationale of the present thesis given the critical mismatch between human readers performance and DRC model simulations that has been documented. In particular, whereas SQ and N are shown to exert additive effects on human RTs, they interact in the DRC model simulations, with the effect of SQ smaller for high-N stimuli.

In the DRC model, this result is caused by the interactive activation between the letter level and the orthographic lexicon. As said, degradation is implemented by reducing the weights of the connections between the feature and the letter levels; as a consequence, reducing SQ slows down the rate at which activation accrues at the letter level. When the nonword in input has orthographic neighbours, the corresponding lexical entries will be activated in the orthographic lexicon and, in turn, activation feeds back to the letter level. In other words, the lexical entries corresponding to the orthographic neighbours of the nonword in input send activation back to the target letters, partially compensating for the effect of degradation. The delay in processing due to the reduction of SQ would be thus reduced as the number of orthographic neighbours of a nonword increases.

In order to make the DRC model able to reproduce the pattern of observed data, Reynolds and Besner (2004) suggested to add a threshold at the letter level: a threshold at this level would in fact prevent the interactive activation between the letter level and the orthographic lexicon thus rendering the effects of SQ and N additive. We argue, however, that a threshold at the letter level is not a plausible account in this context. In order to explain the additive effects of SQ and N, in fact, one need to assume that the output of the letter level is thresholded before it activates the lexical route. This hypothesis is clearly inconsistent with previous empirical data, such as the interaction between SQ and repetition (Blais & Besner, 2007) and the interaction between SQ and semantic priming (Ferguson et al., 2009) that have been obtained when reading words.

The aim of the present study is to further test the hypothesis of a threshold in the reading system. In particular, the joint effects of SQ and N have been analyzed by using a particular manipulation of this latter variable. In the experiment reported by Reynolds and Besner (2004) the variable N has been manipulated by varying the number of orthographic neighbours, i.e. nonwords with few neighbours (low-N nonwords: mean = 2.95, SD = 1.31) and nonword with many neighbours (high-N nonwords: mean = 12.75, SD = 2.30) have been presented in the task. However, both the two types of stimuli would produce interactive activation between the orthographic lexicon and the letter level in the DRC model (even if in a theoretically different amount). It follows that, according to the model, the effect of degradation would be reduced both for the low-N nonwords and for the high-N nonwords used in this task. As a consequence, this may be not the adequate manipulation to analyze the joint effects of SQ and N in nonword reading.

We argue that the effects of SQ and N can be analyzed in order to determine whether the presence/absence of orthographic neighbours makes any difference on stimulus degradation. In other words, the question here is whether having orthographic neighbours would facilitate degraded nonword reading compared to the condition in which nonwords have no neighbours. The prevision of the DRC model we want to test is that lexical knowledge should reduce the effect of degradation in nonword reading. To this end, the joint effects of SQ and N have been analyzed in a reading aloud task on English skilled readers and in DRC model simulations by presenting zero-N nonwords and many-N nonwords (matched in terms of Total Letter Confusability) either in clear or in degraded conditions. According to the DRC model, an interaction between SQ and N should be obtained, with the effect of SQ being smaller for many-N nonwords than for zero-N nonwords.

The ultimate goal of the present experiment is to test the hypothesis of a threshold: if the letter level is thresholded, then the effects of SQ and N should be additive in this experiment. In fact, a threshold at the letter level would prevent the interactive activation between the letter level and the orthographic lexicon and the presence/absence of orthographic neighbours should not play any role on stimulus degradation. In addition, further evidence regarding the locus of the N effect in reading will be provided; the interpretation proposed by Reynolds and Besner (2004) is in fact valid only assuming that processing at the letter level analysis is thresholded. More in general, the data sustaining a late account of the N effect for skilled readers generally assume that processing in reading occurs in (at least partially) serial stages.

5.2 Method

Participants. Twenty students at the Macquarie University who had English as their first language and normal or corrected-to-normal vision participated as part of their courses requirement.

Design. The experiment consisted of a 2x2 within-subjects design with N (zero-N vs. many-N nonwords) and stimulus quality (SQ; clear vs. degraded conditions) as factors.

Material. A total of 160 orthographically legal monosyllabic nonwords with five letters in length have been used (these stimuli can be seen in the Appendix, section D). Nonwords have been derived from the ARC Nonwords Database (Rastle, Harrington, & Coltheart, 2002) so to avoid pseudohomophones. The items belong to two groups of 80 stimuli representing the zero-N condition and the many-N condition. Zero-N nonwords were nonwords without any orthographic neighbour, whereas high-N nonwords had an average of 7.75 orthographic neighbours (range: 7-10; neighbourhood mean frequency: 36.9 occurrences per million). Zero-N and many-N nonwords were balanced in terms of TLC (2.38 vs. 2.38, $t<1$; LC values have been determined by averaging empirical letter-confusion matrices for upper-case letters; Gilmore et al., 1979; Loomis, 1982; Townsend, 1971; Van Der Heijden, et al., 1984) and number of whammies (1.01 vs. 1.06, $t<1$; see Rastle & Coltheart, 1998). At each of the two levels of N, two lists of 40 nonwords have been created (list A and list B) in order to assign half the items to the clear condition and the other half to the degraded condition. The many-N nonwords were balanced for N (7.75 vs. 7.75, $t<1$), neighbourhood mean frequency (35.7 vs. 38.2, $t<1$), TLC (2.38 vs. 2.38, $t<1$), and number of whammies (1.1 vs. 1; $t = 1.1$, n.s.) across these two lists. The zero-N nonwords were balanced in terms of TLC (2.38 vs. 2.38, $t<1$) and number of whammies (1 vs. 1; $t <1$) across the two lists. Finally, the initial phoneme was matched in the four cells. Each participant saw 80 stimuli clear (40 zero-N nonwords and 40 many-N nonwords) and 80 stimuli degraded (40 zero-N nonwords and 40 many-N nonwords). The assignment of stimuli to the four conditions was counterbalanced across participants, in such a way that half the subjects saw the items of the list A under the clear condition and the items of the list B under the degraded condition whereas the remaining subjects saw the items of the list A under the degraded condition and the items of the list B under the clear condition.

Apparatus. The experiment took place in a sound attenuated and dim lit room. Stimuli presentation and data recording were controlled by DMDX software and running on a personal computer. RTs and errors were determined by using CheckVocal software. The display was synchronized with the screen refresh cycle. Stimuli were presented centrally in upper-case letters in the 18-point Courier New font on a black background. Clear stimuli were displayed in white (RGB values: 85,85,85); degraded stimuli were displayed in grey (RGB values: 5,5,5). Responses were collected via a microphone connected to a voice key assembly. Response latencies were timed from stimulus onset to voice key activation, which also terminated the display.

Procedure. Participants were tested individually and sat in front of a computer screen. They were instructed to read each letter string aloud as quickly as possible and to minimize errors. A 18 items practice session preceded the experimental session. Each trial began with a 500 ms presentation of a fixation point at the centre of the computer screen followed by a 200 ms presentation of a blank. Immediately after the stimulus appeared and remained on the screen until a response was registered by the voice key or 3 sec elapsed. Stimuli were presented in a random order for each participant. Responses were coded offline as correct or incorrect by the experimenter using CheckVocal software. The experimenter determined RTs using the waveform recorded by this software in order to reduce measurement error associated with voice key timing and correct for mistrial (i.e., voice key failure).

5.3 Results

Pronunciation errors (11,3%) were removed prior to reaction times analysis. Correct reaction times were submitted to the Van Selst and Jolicoeur's (1994) trimming procedure. Outliers (1.5%) were removed prior to RTs analysis. Mean RTs according to conditions and percentages of error are reported in Table 5.

	Stimulus quality					
	Clear		Degraded		Diff.	
Orthographic Neighborhood	RT	E%	RT	E%	RT	E%
Zero-N	717	12	993	17	276	5
Many-N	659	6	892	11	233	5
Diff.	58	6	101	6		

Table 5. Mean reaction times (RTs) and percentages of error (%E) according to conditions.

In the ANOVA for the participants (F1) N and SQ were repeated factors. In the ANOVA for items (F2) N was an independent factor and SQ was a repeated factor.

RTs. Analysis showed a main effect of N, $F1(1, 19) = 89.3$, $MSE = 1416$, $p < .001$, $F2(1, 158) = 29.7$, $MSE = 20663$, $p < .001$, and a main effect of SQ, $F1(1, 19) = 113.6$, $MSE = 11395$, $p <$

.001, $F2(1, 158) = 1043$, $MSE = 5126$, $p < .001$. Crucially, the two effects were qualified by a significant interaction, $F1(1, 19) = 13.6$, $MSE = 9093$, $p < .005$, $F2(1, 158) = 8.1$, $MSE = 5126$, $p = .005$, with the effect of degradation larger for the zero-N nonwords than for the many-N nonwords.

Accuracy. Analysis showed a main effect of N, $F1(1, 19) = 21.1$, $MSE = .003$, $p < .001$, $F2(1, 158) = 11.3$, $MSE = .023$, $p = .001$, and a main effect of SQ, $F1(1, 19) = 13.1$, $MSE = .004$, $p < .005$, $F2(1, 158) = 18.4$, $MSE = .010$, $p < .001$. However, the interaction between the two factors was not significant, $Fs<1$.

5.4 Simulation

The set of nonwords used with participants was run through two different computational versions of the DRC model.

In the first simulation the version of the DRC model presented in the *Psychological Review* 2001 (Coltheart et al., 2001) – that we call DRC-PR – has been used. This version of the model has been employed since the simulations reported by Reynolds and Besner (2002; 2004) have been performed by using the DRC-PR.

In the second simulation the currently public version of the DRC model – the DRC 1.2 – has been used[28].

5.4.1 DRC-PR

The items were run through the DRC-PR under both the clear and the degraded condition. In the degraded condition the weights between features and letters were reduced by 40% (as in Reynolds & Besner, 2004); specifically, the feature-to-letter excitation parameter was reduced to .003, whereas the feature-to-letter inhibition parameter was reduced to .09.

The model made 9 lexicalization errors (CRAME, CRASE, CROSE, PROPE, PROME, DRAVE, FLATE, SLIPE, CRYBE) and therefore these items were discarded from the analyses.

[28] Both the versions of the DRC model are downloadable from the DRC web site at http://www.maccs.mq.edu.au/~ssaunder/DRC/. The differences between the two versions are also documented on the web site. Following the principle of nested modelling, the DRC 1.2 has been tested to ensure that is capable of reproducing all the effects that the DRC-PR could simulate; moreover, there are data from experiments with Masked Onset Priming which can be simulated by the DRC 1.2 but not by the DRC-PR (see Mousikou, Coltheart, Saunders, & Yen, 2010).

Mean cycles to criterion are reported in Table 6.

	Stimulus quality		
	Clear	Degraded	Diff.
Orthographic Neighborhood	Cycles	Cycles	Cycles
Zero-N	168	182	14
Many-N	158	170	12
Diff.	10	12	

Table 6. Mean cycles according to conditions.

An ANOVA with SQ as repeated factor and N as independent factor was conducted on cycles. The DRC-PR behaviour mimed that of humans. Analysis showed a main effect of N, $F(1, 149) = 16.7$, $MSE = 494.4$, $p < .001$, a main effect of SQ, $F(1, 149) = 1773.9$, $MSE = 7.4$, $p < .001$, and a significant interaction, $F(1, 149) = 53.7$, $MSE = 7.4$, $p < .01$, imputable to the size of the SQ effect being smaller for the items with many-N than for the items with zero-N.

5.4.2 DRC 1.2

The items were run through the DRC 1.2 under both the clear and the degraded condition. As in the previous simulation, the weights between features and letters have been reduced by 40% in the degraded condition.

The parameter set of the DRC 1.2 has been modified in order to allow the model to correctly simulate the neighbourhood size effects in nonword reading. In particular, the parameter that regulates inhibition from letters to words is currently very high in the DRC 1.2 (specifically, it is set to .48). The highest this parameter is, the less a letter string can excite potentially supportive neighbours; as a consequence, with the current value, even when the input letter string is just a single letter different from some real word, the entry in the orthographic lexicon for that word won't be activated. The consequence is the inability of the model to correctly simulate the effects due to the orthographic neighbourhood in reading. Hence, the value of the letter-to-word inhibition parameter has been reduced both in the clear and in the degraded condition in order to allow the model to correctly simulate these effects[29].

[29] The Letter-to-OrthographicLexicon-Inhibition parameter has been set to .435, which is the default value used in the DRC-PR. Coltheart et al. (2001) suggested to further reduce this parameter to .350 in order to correctly simulate the

The model did not make any error. Mean cycles to criterion are reported in Table 7.

	Stimulus quality		
	Clear	Degraded	Diff.
Orthographic Neighborhood	Cycles	Cycles	Cycles
Zero-N	138	165	27
Many-N	132	157	25
Diff.	6	8	

Table 7. Mean cycles according to conditions.

An ANOVA with SQ as repeated factor and N as independent factor was conducted on cycles. The DRC 1.2 behaviour mimed that of humans. Analysis showed a main effect of N, $F(1, 158) = 72.9$, $MSE = 51$, $p < .001$, a main effect of SQ, $F(1, 158) = 8883.5$, $MSE = 6.2$, $p < .001$, and a significant interaction, $F(1, 158) = 14.4$, $MSE = 6.2$, $p < .001$: the size of the effect of SQ is smaller for many-N nonwords than for zero-N nonwords.

5.5 Discussion

The experiment evidenced three different results, all correctly reproduced in the simulations of the DRC model (both the DRC-PR and the DRC 1.2 versions).

First, nonwords with neighbours are named faster than nonwords without neighbours. This result is known as the N effect (McCann & Besner, 1987). Both the Italian (see Chapter 2) and the English versions of the DRC model correctly simulate this effect.

Second, we found an effect of SQ: clear stimuli are read aloud faster than degraded stimuli. This effect is correctly simulated by the DRC model when the strength of the connections between the feature and the letter units is reduced to simulate the degraded condition. Specifically, a reduction of SQ has been simulated in the DRC model by implementing the manipulation used by

effects of N in word reading. However, we didn't use this value since it would have produced a very poor performance in reading nonwords. Moreover, it has been shown that the value .350 is indeed necessary to allow the DRC model to correctly simulate the effect of the phonological and orthographic neighborhood in word reading (see Mulatti, Reynolds, & Besner, 2006); however, different phonological and orthographic neighborhood effects have been obtained in nonword reading (see Reynolds, Mulatti, & Besner, 2006) and this further reduction is not required for these simulations.

Reynolds and Besner (2004), i.e. a reduction of the weights of the connections between the feature and the letter units by 40%.

Third, the size of the SQ effect depends upon the size of N: the effect of degradation is smaller for nonwords with many-N than for nonwords without orthographic neighbours. In other words, zero-N nonwords are harmed by stimulus degradation more than many-N nonwords. This result is explained within the dual-route account by the interactive activation between the letter level and the orthographic lexicon. When a nonword with orthographic neighbours is presented to the system, its neighbours will be activated in the lexicon and, in turn, activation will spread to the later stages. At the same time, the orthographic lexicon will activate the letter level via feedback connections, thus contributing activation to the target letters and partially compensating for the delay produced by stimulus degradation. On the contrary, there is no feedback from the lexicon when the nonword in input has no orthographic neighbours and the effect of degradation is thus stronger for these stimuli. Clearly, such an interaction cannot be explained by assuming a threshold at the letter level. A threshold at the letter level would in fact prevent the interactive activation between this level and the orthographic lexicon; hence, the interaction between SQ and N (zero-N vs. many-N) would be eliminated and additive effects of the two variables would be expected.

Our results may have important implications for the locus of the N effect in reading. The interaction between SQ and N obtained in our experiment is in fact consistent with an early locus, thus indicating that at least part of the N effect for skilled readers arises through the interactive activation between the orthographic lexical units and the letter units. Nevertheless, our results are also consistent with a late account. An interaction between SQ and N would in fact be incompatible with a late locus of the N effect only assuming a (at least partially) thresholded system in which the effect due to degradation is resolved early in processing. According to Reynolds and Besner (2004) this would be the case given the additive effects of SQ and nonword letter length in reading (Besner & Roberts, 2003) and the additive effects of SQ and word frequency in lexical decision tasks (e.g., Balota & Abrams, 1995; Borowsky & Besner, 1997); these results would in fact demonstrate that the effect of degradation is resolved prior to phonological processing and, indeed, prior to the effect of word frequency. Conversely, we showed in Chapter 4 that SQ and letter string length interact in nonword reading under the appropriate experimental conditions; moreover, as we will discuss in details in the next chapter, interactions between SQ and word frequency have been also documented in reading (see O’Malley & Besner, 2008; Yap & Balota, 2007). Hence, it seems to us that there is not convincing evidence indicating that the effect of degradation is resolved early in the reading system; more likely, SQ influences processing downstream, according to the cascaded assumption. If we adopt a model in which processing operates in a cascaded fashion, factors that interact with one another could have their effects at the same level of processing as well as be influencing different processes; as a consequence, an interaction between SQ and N would not uniquely suggest that N has its effect early in the reading system but this result will be also consistent with a later effect.

In general, we suggest that the early-locus and the late-locus accounts are not alterative hypotheses. Lexical knowledge may in fact influence skilled readers when reading nonwords both through the interactive activation between the letter units and the orthographic lexicon and through the feed-forward connections from the orthographic lexicon to the phoneme system. This interpretation is also consistent with the simulation of the N effect within the DRC framework (see Reynolds & Besner, 2002).

Clearly, an issue remains however to be explained. Why additive effects of SQ and N (low-N vs. high-N) are obtained by Reynolds and Besner (2004), whereas an interaction between SQ and N (zero-N vs. many-N) is found in our experiment?

It could be possible that these apparently inconsistent results depend on the different manipulation of the orthographic neighbourhood size. In fact, the effect of N is likely to be non-linear, i.e. the larger N is, the smaller the effect of increasing the number of orthographic neighbours will be. This means that the biggest N effect may be obtained when N is manipulated between 0 and 1, the next biggest effect when N is manipulated between 1 and 2, and so on. If that is so, then it will be best always to include a N = 0 condition in experiments manipulating this variable. Moreover, it might be possible that even the low-N nonwords in the experiment of Reynolds and Besner (2004) provided a large enough amount of feedback from the orthographic lexicon to the letter level to help counter the difficulties in letter identification caused by degradation. Hence, the effect due to the reduction of stimulus quality may be reduced both for the low-N nonwords and for the high-N nonwords in their study. It might be therefore possible that what really makes a difference on the effect due to stimulus degradation is having or not orthographic neighbours, rather than a difference in their amount.

Nevertheless, the DRC model still fails to simulate the results obtained with human readers. In fact, since the DRC model is suppose to mime the human performance, then it should reproduce the additivity between SQ and N (low-N vs. many-N) that has been observed. In other words, if different effects between SQ and N are obtained depending on the particular manipulation of the variable N, then the DRC model should be able to simulate the whole pattern of results caused by these manipulations. Clearly, further work is needed in this context in order to investigate these issues.

An explanation that appears to us being promising is by considering the Total Letter Confusability (TLC), a variable that – as demonstrated in Chapter 3 – influences nonwords reading when the stimuli are degraded in the task. In particular, we argue that the additive effects between SQ and N obtained by Reynolds and Besner (2004) might be due to a confounding with this variable and that when this confounding is removed the true result could be an interaction, as predicted by the DRC model.

In the previous chapter we demonstrated a confounding with TLC in the study involving the manipulation of letter string length in degraded nonword reading. However, while the relationship between length and TLC was clearly reasonable (i.e., since longer nonwords have more letters, one might expect that the TLC is higher for long nonwords than for short nonwords), the role of this variable is not obvious in this context. Nevertheless, an analysis performed on the stimuli used in the Reynolds and Besner's (2004) experiment indicates that the TLC was significantly higher for the high-N nonwords than for the low-N nonwords used in this study, 410 vs. 406.7; $t(82) = 2.326$, $p = .022$[30].

The hypothesis of a confounding between N and TLC is thus plausible. The effects of SQ and N might result additive in the Reynolds and Besner's (2004) study because the partial compensation of the SQ effect due to the larger amount of orthographic neighbours was masked by the higher confusability values associated to the high-N nonwords. In other words, since the high-N nonwords had higher TLC values than the low-N nonwords used in this study, the former stimuli might be delayed more in degraded presentation; hence, the true interaction between SQ and N would not emerge. We argue that if the high-N and the low-N nonwords would be matched on the TLC, then the effect of degradation could be smaller for high-N nonwords than for low-N nonwords and the true result could be thus the interaction predicted by the DRC model.

Further experiments with skilled readers are clearly called for. For example, one way to determine the plausibility of our hypothesis is by running an experiment in which the effects of SQ and N are analyzed when high-N and low-N nonwords are matched for TLC. If our hypothesis is plausible, we expect an interaction between SQ and N (low-N vs. high-N) in this experimental condition. Instead, the additive effects obtained by Reynolds and Besner (2004) should be replicated when the high-N nonwords are chosen so to have higher confusability values than the low-N nonwords used in the task.

To conclude, the studies analyzing the joint effects of SQ and N (low-N vs. high-N) showed a critical mismatch between the DRC model simulations and the human readers performance. A threshold at the letter level has been proposed in order to eliminate this mismatch. Our experiment

[30] The Reynolds and Besner's (2004) experiment used lowercase letters. LC has been thus derived from the Courrieu et al.'s (2004) confusion matrix for lowercase letters.

clearly shows that a threshold at the letter level is an inadequate solution since SQ and N interact when the presence/absence of nonwords' orthographic neighbours is jointly manipulated with stimulus degradation in the reading task. Despite other explanations are possible, we argue that a confounding with the TLC might explain the additive effects that have been previously observed. Clearly, further empirical investigation is needed in this context in order to define whether and how the different results obtained in these studies are caused by the TLC or rather depend on the different manipulation of the variable N.

Moreover, another interesting issue might be to analyze the effect of N size and degradation in word reading. In fact, at least two hypotheses can be formulated here. From one hand, SQ and N might interact, with the N effect being larger for degraded than for clear stimuli; following a similar argument to the one previously described for nonwords, the feedback from the lexicon may in fact help target letter recognition in word reading, thus reducing the effect of degradation more strongly when N is high than when it is low. From the other hand, the interaction between N and SQ might be reverse, with the N effect being larger in the clear than in degraded condition. The effect of degradation in word reading may in fact be stronger as N increases, since degrading the letters of a word that has many neighbors could enhance the level of uncertainty about the correct answer and thus increases the response latency.

6 TOTAL LETTER CONFUSABILITY AND LEXICALITY IN DEGRADED READING

Inconsistent results with the DRC model's cascaded assumption have been obtained also in experiments involving factorial manipulations of stimulus degradation and a lexical variable such as word frequency and lexicality.

The joint effects of SQ and word frequency have been largely analyzed in lexical decision, where the two effects have been reported to be additive (e.g., Balota & Abrams, 1995; Borowsky & Besner, 1993; O'Malley et al., 2007; Plourde & Besner, 1997; Yap & Balota, 2007; Yap, Balota, Tse, & Besner, 2008). However, O'Malley et al. (2007) and Yap and Balota (2007) also reported interaction of stimulus quality and word frequency in reading aloud. O'Malley and Besner (2008) concluded that the different results obtained in these tasks may be due to the presence/absence of nonwords; more specifically, the two factors would interact when only words are presented in the task (e.g., reading aloud), whereas they would exert additive effects when nonwords are part of the stimulus set (e.g., lexical decision). This interpretation has been confirmed in further reading aloud experiments. O'Malley and Besner (2008) showed, in fact, that when only words were presented in the reading task, the effects of SQ and word frequency interacted, with the effect of SQ being larger for low-frequency words than for high-frequency words. However, when also nonwords were included within the experimental stimuli, word frequency and SQ exerted additive effects on reading latencies. Importantly, the effect of lexicality has also been analyzed when jointly manipulated with stimulus contrast: the two variables have been reported to exert additive effects on skilled readers latencies in several different studies (Besner & O'Malley, 2009; Besner et al., 2010; O'Malley & Besner, 2008); of course, this manipulation also implies the presence of both words and nonwords in the task.

Critically, the additivities of SQ and a lexical variable (i.e., word frequency and lexicality) are inconsistent with models operating in a purely cascaded fashion. In order to explain these results O'Malley and Besner (2008) proposed that readers might switch from cascaded processing to thresholded processing as a function of the experimental context (i.e., lexicalization hypothesis). In particular, they suggested that the letter level would be thresholded when both words and nonwords are presented in the reading task whereas, when the list is composed of words only, processing would flow in a purely cascaded fashion. The present study aims to test this hypothesis.

6.1 Introduction

Computational models of reading aloud and visual word recognition usually implement either cascaded processing – as in the DRC model (Coltheart et al., 2001) – or thresholded processing – as in the *logogen* model (Morton, 1969). However, O'Malley and Besner (2008) recently proposed that readers might switch from cascaded processing to thresholded processing as a function of the experimental context.

O'Malley and Besner (2008) called their proposal the "lexicalization hypothesis". This account specifically states as follow: when the experimental list of a reading aloud task includes both words and nonwords, participants would threshold the output of the letter level to prevent lexical capture of nonwords; when the list is composed of words only, processing would flow in a purely cascaded fashion. To note, nonwords with orthographic neighbours are implicated here.

In a cascaded model as the DRC, lexical capture might in fact occur when the stimulus is degraded: a nonword may activate a word form representation enough to be erroneously read as the word instead of the nonword. O'Malley and Besner argued that a threshold at the letter level would prevent (or, at least, reduce the frequency of) lexical captures: only once the letters comprising the stimulus are unambiguously identified will activation be passed on to later stages, and so the possibility of erroneously selecting a word given a nonword is reduced or eliminated.

O'Malley and Besner (2008) developed their lexicalization hypothesis to explain data they obtained in reading aloud experiments. In these experiments, when the experiment's stimuli consisted solely of words, the variables SQ and word frequency (high-frequency vs. low-frequency) interacted: the effect of stimulus quality was larger when frequency was lower. This result is not new (O'Malley et al., 2007; Yap & Balota, 2007) and is predicted by cascaded models such as the DRC (see Reynolds & Besner, 2004, for simulations). However, O'Malley and Besner (2008) also found that when an experiment's stimuli included both words and nonwords (with orthographic neighbours), word frequency and SQ had additive effects on reading aloud latencies. Moreover, additive effects of SQ and lexicality (words vs. nonwords) have been largely documented in literature (Besner & O'Malley, 2009; Besner et al., 2010; O'Malley & Besner, 2008). These additivities are inconsistent with models operating in a purely cascaded fashion. In contrast, if we assume, as O'Malley and Besner suggest, that when the stimuli consist of both words and nonwords the letter level is thresholded (to avoid lexical captures of nonwords), then the result is easily accounted for: the idea is that since SQ affects the perceptual levels and lexicality and word frequency affect the subsequent lexical level, thresholding the letter level prevents interaction between those two stages, rendering the effects of SQ and lexical-level variables (i.e., lexicality and word frequency) additive.

We argue that a threshold at the letter level might be not an adequate solution. In fact, reducing letter contrast has its effect at the visual feature analysis level because a visual feature will take longer to resolve when its contrast with the background is low. It seems in fact unarguable that the processing level at which SQ has a direct effect is not the letter level but the (earlier) visual feature level.

If one wanted to investigate whether there is a threshold specifically at the letter level (as it has been suggested) then one should investigate whether a variable whose direct effect is at that level has additive effects with a lexical variable such as lexicality or word frequency. We argue that a similar variable could be the Total Letter Confusability (or TLC): increasing the Total Letter Confusability has its effect at the letter level because when a letter has high confusability it will activate other letters – the confusable ones – at the letter level, introducing competition between letters which will slow the rise of activation of the correct letter at that level.

The aim of the work reported here was to test the lexicalization hypothesis. This hypothesis makes the general prediction that when both words and nonwords are present in a reading aloud task and stimuli are degraded (i.e., lexical capture may occur), the letter level will always be thresholded, and so any variable which affects the operation of the letter level will have additive effects with any variable that affects the operation of the orthographic-lexicon level. With a reading aloud task we orthogonally manipulated a variable that would have an effect at the letter level (i.e., the TLC) and a variable that would have an effect at the orthographic-lexical level (i.e., lexicality: word vs. nonword). There were both words and nonwords in our experiment and all the stimuli have been presented in reduced contrast, so all the conditions postulated by the lexicality hypothesis are met. The lexicalization hypothesis thus predicts that TLC and lexicality will have additive effects on reading aloud latencies.

Before proceeding an issue has to be analyzed. If words and nonwords are not matched on TLC, an apparent interaction between degradation and lexicality might instead be an interaction between degradation and letter confusability, which would require a completely different theoretical interpretation. It is important therefore to determine whether either of the stimulus variables manipulated by O'Malley and Besner (2008) – namely, word frequency and lexicality – were confounded with TLC. Their experiments used lowercase letters and the Courrieu et al.'s (2004) confusion matrix for lowercase letters has thus been used; our analyses of the materials used by O'Malley and Besner (2008) indicated that the high-frequency and low-frequency words they used in their Experiments 1 and 2 did no differ in TLC, and nor did the high-frequency and low-frequency words they used in their Experiment 3. We also found that there were no differences in TLC between words and nonwords in any of their experiments. So their pattern of results did not arise because of any confounding with TLC in their materials.

Nevertheless, it remains the case that according to the lexicalization hypothesis, when a stimulus list comprises words and nonwords, the letter level gets thresholded. If the letter level is thresholded, it follows that TLC should affect both words and nonwords equally, and so the lexicalization hypothesis predicts additive effects of TLC and lexicality in these circumstances. An interaction of TLC with lexicality would be hence evidence against the lexicalization hypothesis.

6.2 Method

Participants. Twenty students at the Università degli Studi di Padova who had Italian as their first language and normal or corrected-to-normal vision participated as volunteers.

Design. The experiment consisted of a 2x2 within-subject design with lexicality (words vs. nonwords) and Total Letter Confusability (TLC; low TLC vs. high TLC) as factors.

Materials. A set of 160 upper-case stimuli with five letters in length was selected (these stimuli can be seen in the Appendix, section E). This set was composed of 80 disyllabic Italian words and 80 disyllabic nonwords. Words were all of low written frequency (mean: 16.7 occurrences per million). The nonwords were all pronounceable and were derived from words by changing one letter provided the initial phoneme of that word remained intact; the words used to derive the nonwords had a mean written frequency similar to that of the words used as stimuli, namely 18.5 occurrences per million. Letter confusability was determined from empirical letter-confusion matrices obtained in previous studies (Gilmore et al., 1979; Loomis, 1982; Townsend, 1971; Van Der Heijden et al., 1984). TLC was obtained by summing the confusabilities of individual letters in the string. Forty words were classified as having high TLC (mean: 2.6) and forty as having low TLC (mean: 1.9; $t(78) = 23$, $p < .001$). High and low TLC words were balanced in terms of written frequency (16.9 vs. 16.5; $t<1$), length in number of letters, and neighbourhood size (6.5 vs. 6.5). Forty nonwords were classified as having high TLC (mean: 2.6) and forty as having low TLC (mean: 1.9; $t(78) = 25$, $p < .001$). High and low TLC nonwords were balanced in terms of baseword frequency (19.1 vs. 17.9; $t<1$), neighbourhood size (4.4 vs. 4.4; $t<1$), length in number of letters, and the position of the letter changed to derive the nonword from the baseword (3 vs. 3; t<1; see Mulatti et al., 2007). The phonological onsets of the stimuli were matched across conditions.

Apparatus. The experiment took place in a sound attenuated and dimly lit room. Stimuli presentation and data recording were controlled by software developed in E-prime and running on a personal computer. The display was synchronized with the screen refresh cycle. Stimuli were presented centrally in upper-case letters on a black background (RGB values: 0, 0, 0). The stimuli were displayed in the 18-point Courier New font. All the stimuli were displayed in degraded mode (RGB values: 8, 8, 7). Responses were collected via a microphone connected to a voice-key assembly. Response latency was timed from stimulus onset to voice key activation, which also terminated the display.

Procedure. Participants were tested individually. They sat in front of the computer screen and the microphone was placed directly in front of but slightly below the subjects' face, so as not to obstruct screen view. Participants were instructed that when a letter string appeared on the screen, their task was to pronounce it as quickly and accurately as possible. They were informed that the stimuli could be either a word or a nonword. Subjects were then presented with 12 practice trials. The 160 experimental stimuli followed the practice session after a short pause. Each trial began with a 500 ms presentation of a fixation point at the centre of the computer screen followed by a 200 ms presentation of a blank screen. Immediately after the stimulus appeared and remained on the screen until a response was registered by the voice key or 3 sec elapsed. The inter-trial-interval was set to 2 sec. Stimuli were presented in a random order, i.e. the variables of Lexicality and Total Letter Confusability were randomized, not blocked. The experimenter coded the pronunciation as correct on the basis of the standard set of Italian grapheme-phoneme rules, voice key triggering failure, lexicalization error or articulation error.

6.3 Results

Correct reaction times were submitted to the Van Selst and Jolicoeur's (1994) outlier removal procedure. Outliers (1.5 %) were removed prior to reaction times analysis. Mean reaction times and percentages of error are reported in Table 8.

	Words		Nonwords		Diff.	
TLC	RT	%E	RT	%E	RT	%E
High	713	6.4	814	10.9	101	4.5
Low	709	8.2	765	10.2	56	2.0
Diff.	4	-1.8	49	.7		

Table 8. Reaction times (RTs) and percentages of error (%E) according to conditions.

ANOVAs were conducted for reaction times and errors. For the participant analysis (F1, t1), Lexicality (words vs. nonwords) and TLC (high vs. low) were treated as repeated factors. For the item analysis (F2, t2), Lexicality and TLC were treated as independent factors.

RTs. The analysis revealed a main effect of Lexicality, $F1(1, 19) = 24.7$, $MSE = 4940$, $p < .001$, $F2(1, 156) = 57.3$, $MSE = 3951$, $p < .001$, and a main effect of TLC, $F1(1, 19) = 13.1$, $MSE = 1060$, $p < .005$, $F2(1, 156) = 6.3$, $MSE = 3951$, $p < .05$. However, the main effects were qualified by a significant interaction, $F1(1, 19) = 12.7$, $MSE = 798$, $p < .005$, $F2(1, 156) = 6.7$, $MSE = 3951$, $p < .001$. Paired comparisons revealed that whereas TLC affects nonwords so that low TLC nonwords are read faster than high TLC nonwords, $t1(19) = 4.4$, $p < .001$, $t2(78) = 3.3$, $p < .005$, TLC does not affect word reading, $ts < 1$.

Accuracy. The main effect of Lexicality proved significant by participants but not by items, $F1(1, 19) = 14.3$, $MSE = .002$, $p < .005$, $F2(1, 156) = 2.9$, $MSE = .015$, $p = .091$. Neither the main effect of TLC, $Fs < 1$, nor the interaction, $F1(1, 19) = 1.1$, $MSE = .003$, $p > .3$, $F2 < 1$, were significant.

6.4 Discussion

Our results demonstrated that whereas TLC affects nonword reading, it does not affect word reading. In other words, a variable having its effect at the letter level interacts with a variable that has an effect at the lexical level. This result contradicts the lexicalization hypothesis which predicts additivity of these factors when both words and nonwords are present in the experiment and stimuli are degraded. This means that the letter level is not always thresholded when the criterion for the presence of a threshold postulated by the lexicalization hypothesis is met.

However, an issue remains to be explained. Why one factor which affects the recognizability of letters (i.e., SQ, manipulated by varying letter contrast) has additive effects with a variable that operates at the lexical level, whereas another factor which also affects the recognizability of letters (i.e., Total Letter Confusability) interacts with a variable that operates at the lexical level?

We propose that this is because although both factors influence letter recognizability, they have their effects at different levels. Reducing letter contrast has its effect at the visual feature analysis level because a visual feature will take longer to be identified when its contrast with the background is low. Instead, increasing the TLC has its effect at the subsequent letter identification level because when a letter has high confusability it will activate other letters that will compete with the correct letter within the letter level.

The interaction of TLC and lexicality is easily interpreted within an IA framework. When a letter is highly confusable with other letters, it will activate at the letter level the representations of these other letters (as well as its own representation), and so there will be competition at the letter level, which will slow processing. If the stimulus is a word, the target letter receives both feed-forward activation from the visual features of the stimulus and feedback activation from the lexical representation of the stimulus. The feedback from the lexicon assists target letter recognition by contributing activation to the letter detectors for the correct letters and inhibition to the competing letter detectors, and so could compensate for the interference produced by the competing non-target letters. If the stimulus is a nonword, there is no feedback from the lexicon: target letters receive activation from the visual feature level only and thus suffer more from the activation of the competing, similar non-target letters. Given that the feedback from the lexical level to the letter level will be stronger for high-frequency than for low-frequency words, there should also be an interaction between word frequency and TLC, a prediction worth testing.

What about the additivity of a variable affecting the visual feature level (i.e., stimulus contrast) and a variable affecting the lexical level (i.e., word frequency or lexicality)?

A possible explanation occurs by considering the effects due to list composition. In fact, in the experiments presented by O'Malley and Besner (2008), stimulus quality and word frequency interacted when only words were presented, with the effect of stimulus quality being smaller for high-frequency words (see also Yap & Balota, 2007). However, the two variables exerted additive effects on reading latencies when words and nonwords were mixed together in the task. Moreover, the additive effects obtained between SQ and lexicality also imply the presence of both words and nonwords in the task. List composition is thus the crucial variability in these experiments.

The route emphasis account has been proposed in order to explain the effect of list composition in the context of the dual-route framework (e.g., Coltheart & Rastle, 1994). The basic idea of the route emphasis account is that readers could strategically adjust the extent to which pronunciation performance relies on the lexical and on the non-lexical routes as a function of the type of stimuli presented in the task. For instance, when only nonwords are present in the task, one might expect more emphasis on the non-lexical route and an attenuation of the lexical route; conversely, when only words are present, the non-lexical route would be de-emphasized, increasing the reliance of the lexical route.

Our hypothesis is based on this idea. We propose that the balance between the two routes could favour the lexical route over the non-lexical route more strongly when only words are present than when words are mixed with nonwords. A similar strategy could be justified by the fact that lexical capture of nonwords would be reduced when the lexical route is de-emphasized.

To date, in DRC model there is not feedback from the letter level to the feature level. This trait was inherit from the McClelland and Rumelhart's (1981) IA model and is justified by the fact that a visual feature can be only turned on by the stimulus in input. Assuming feedback at this level would in fact mean to allow the pattern of features representing the external input to be modified top-down by the activation of functional units at the subsequent levels. Instead, visual features can be clearly activated bottom-up only; in a similar way, there would be no reason to turn a visual feature off once it has been activated. As a consequence, the feedback from the orthographic lexicon does not reach the feature level (where SQ has its effect), but rather the letter level. This means that the effect of lexical variables is only indirect on SQ. In particular, the effect of degradation is transmitted at the letter level given the feed-forward activation form features to letters; moreover, given the interactive activation between the letter level and the orthographic lexicon, the activation from the lexical level feeds back to the letter level: as a consequence the feedback from the lexicon will act on degradation only contributing activation at the letter level.

Our interpretation of the previous results is as follow. When only words are present in the task, the lexical route would be favoured over the non-lexical route more strongly as compared to the condition in which words are mixed with nonwords; hence, the feedback from the lexicon could be fast enough to partially compensate for the effect of degradation (acting at the letter level) in the former condition. SQ and word frequency would thus interact when only words are presented in the task. However, when words and nonwords are mixed together, the lexical route would be weaker and the activation from the lexicon may thus reach the letter level later, without producing any effect on degradation. SQ and word frequency would be thus additive, as well as SQ and lexicality would be additive, when words and nonwords are mixed together in the reading task.

Since TLC has its effect at a different and subsequent level (i.e., the letter level), the feedback from the orthographic lexicon acts directly on this variable; as a consequence, even when the lexical route is slower, the activation from the lexicon would compensate for the delay associated with high confusability. In other words, the emphasis on the functional routes would be less critical when the factorial manipulation involves TLC because the orthographic lexicon feeds activation back to the letter level, which is the process also affected by this factor. As a consequence, TLC interacts with lexical variables (e.g., lexicality) when words and nonwords are mixed together in the task.

It seems to us that the dual-route emphasis account of list composition effects in reading may thus provide a straightforward explanation of the results obtained in reading experiments involving factorial manipulations of stimulus degradation and lexical variables as well as of the results presented in our study.

6.4.1 List composition effects: alternative accounts

To date, an alternative account explaining list composition effects in reading has been proposed and it can be discussed in this context: the time-criterion account (e.g., Chateau & Lupker, 2003; Kinoshita & Lupker, 2003; Lupker, Brown & Colombo, 1997; Taylor & Lupker, 2001).

According to this theory readers establish a time criterion for when articulation should start. The moment in time when participants release the response would be displaced as a function of the difficulty of the material they are exposed to, i.e. late with difficult items, early with easier items. Importantly, the criterion would be set to a position appropriate for the entire block of stimuli in the task and the main effect would be thus an homogenization of the RTs, i.e. when easy and difficult stimuli are mixed together, the response to difficult/slow stimuli would be faster whereas the response to easy/fast stimuli would be slower compared to when easy and difficult stimuli are presented in separated lists. Hence, when nonwords are added in the reading task as O'Malley and Besner (2008) did, one should control for these effects. Specifically to this study, a condition in which only clear and degraded (high-frequency and low-frequency) words are presented in the task is compared with a condition where clear and degraded nonwords have been added. Clear nonwords should not produce any confounding according to the time-criterion account because their RTs are collocated in a position that is more or less intermediate with respect to the other stimuli in the task, i.e. clear nonwords are slower than clear words but faster than degraded words. Adding degraded nonwords could instead constitute a confounding since these stimuli are the slowest in the task. As a consequence, adding degraded nonwords could render the RTs to fast stimuli slower. Hence, high-frequency words (i.e., the fast stimuli in the task) could be delayed by the presence of nonwords more than low-frequency words (i.e., the slow stimuli in the task). The additive effects of SQ and word frequency obtained when words and nonwords are mixed in the task might be due to this delay. Specifically, such an hypothesis would explain the results obtained as due to a similar confounding if degraded high-frequency words would be delayed by the presence of nonwords more than degraded low-frequency words. However, clear stimuli are faster than degraded stimuli and adding nonwords should thus influence, if anything, clear (i.e., not degraded) high-frequency words. This means that if the potential confounding would be removed, the RTs to clear high-frequency words would be, according to the time-criterion, faster than how reported by O'Malley and Besner (2008) when the task consists of both words and nonwords; as a consequence, the frequency effect would result being larger in the clear than in the degraded presentation. Clearly, this interaction would be in the opposite direction to the one reported in the only word condition. Hence, hypothesizing a problem in list composition on the basis of the time-criterion account would not explain the actual pattern of results.

Moreover, an alternative way to interpret list composition extending the time-criterion idea has been recently proposed by Kwantes and Marmurek (2007). The authors suggested to simulate the effects of list composition in the DRC model by manipulating the reading aloud criterion, a mechanism controlling the level of activation that has to be reached in each phonemic slot of the phonemic buffer before articulation can start. When this parameter is set high the criterion is reached later and, therefore, reading latencies are slowed down. This change, however, not only determines the beginning of reading times but also affects the way in which lexical activation contributes to the process of phonological assembly. In the model, when a lexical unit receives activation, it inhibits all the other units within the lexicon and on successive processing cycles, through the mechanism of lateral inhibition, the system gradually converges on a single unit that corresponds to the target stimulus. Therefore, when the response criterion is set low and naming latencies are short, several lexical units (i.e., all units that share some letters and phonemes with the target) are activated, even if not very strongly, and they all contribute to the assembly of target phonology; according to the authors participants would use a General Activation Strategy (or GAS). In contrast, when the response criterion is set high, activation becomes less diffuse and the contribution of single lexical units increases; participants would employ a Specific Activation Strategy (or SAS). Of course, the probability of observing frequency effects increases in this latter condition. It seems to us plausible to assume that the SAS would be used when only words are presented in the task, whereas the GAS would be favoured when words are mixed to nonwords in the type of experiments reported above. In fact, whereas waiting for the identification of the correct lexical unit may be an useful strategy to use when only words are presented, the general activation of entries in the lexicon could facilitate reading more than the specific activation of a single lexical entry when half of the stimuli presented in the task are nonwords. Furthermore, since the probability to observe an effect due to word frequency increases when the SAS is used, it might be hypothesized that, when only words are presented in the task, the frequency effect would be larger in degraded than in clear presentation: since activation would rise more quickly for high-frequency words than for low-frequency words, the effect of degradation would be smaller for the former stimuli. Instead, since the GAS is less sensitive to lexical variables, the amplitude of the effect of degradation might be insensitive to word frequency when words are mixed with nonwords.

To date, this hypothesis is somehow similar to the explanation given by the dual-route emphasis account. Both these hypotheses, in fact, assume the contribution of the lexical route being stronger in the only word condition. Nevertheless, it remains to be defined whether the joint effects of SQ and lexicality as well as the joint effects of TLC and lexicality may be explained by assuming the use of the GAS in mixed lists; for example, it may be suggested that when activation is diffused in the lexicon, the feedback activation could have an effect only on variables directly affecting the letter level. Additional empirical activities will be necessary to further analyze this issue.

6.4.2 Computational modelling

An additional topic regards the simulation of the empirical findings. In the previous sections we concluded that, even if other explanations are perhaps possible, the route emphasis account of list composition effects in reading is a plausible framework to explain the whole pattern of data. The simulations of the results obtained by O'Malley and Besner (2008) might be thus possible by manipulating the strength of the lexical and non-lexical routes of the DRC model.

A first issue is determining how the functional routes of the model should be manipulated in order to simulate human performance. In our theoretical explanation we suggested that the balance between the two routines would favour the lexical route over the non-lexical route more strongly when only words are present in the task than when words are randomly mixed with nonwords. This balancing, however, might be achieved in the DRC model by using several different implementations: for example, one might choose to manipulate the lexical route, either strengthening it in pure condition or weakening it in mixed condition; otherwise, the manipulation may be on the non-lexical route. It seems to us that many different options might in fact be equally adequate in this context. In order to resolve this issue it will be necessary to direct investigate what strategy participants use in these experimental conditions and hence to further analyze the effects due to list composition in degraded reading. This will be the goal of the experiment reported in the next chapter; this study will be useful also to further distinguish between the different accounts proposed to explain list composition effects in reading. More in general, our working assumption is that, if particular effects due to list composition exist in degraded reading, the interpretation of the results obtained in the context of factorial manipulations of SQ and lexical variables should take these effects into account.

Besides the simulation of the effects due to TLC also discussed in the previous chapters, a further problem concerns how SQ is implemented in the DRC model. In fact, whereas SQ affects the feature level analysis, degradation is simulated in the model by reducing the strength of the connections between the feature and the letter units. It follows that the way SQ is actually implemented in the model does not accurately reflect the effects this variable has for humans. The relevance of this issue is crucial in this context since the pattern of empirical findings may in fact depend on the specific level of processing at which SQ (and TLC) has its effect.

We argue that SQ should be implemented in a different way in the model, i.e. by influencing activation of the feature units. As it will be discussed in details in the final chapter of this thesis, this may be realized as follow. At present, a visual feature in the DRC model can be either on or off, i.e. it can take either the value 1 or 0, respectively. The idea might be to allow a visual feature to accumulate activation over time (i.e., taking every intermediate value between 0 and 1) as a function of the quality of the stimulus in input; hence, whereas under normal visibility conditions a feature unit will be fully activated as the stimulus is presented, it will be only partially activated when the stimulus is degraded; because of the cascaded property of the model, less activated units will reduce the rate of activation downstream in the system, thus producing slower responses.

6.4.3 Conclusion

This study aimed to evaluate the lexicalization hypothesis, stating that the letter level might be thresholded in particular experimental conditions (O'Malley & Besner, 2008). The results we obtained falsified this account by showing that a variable having an effect at the letter level (i.e., TLC) interacts with a variable that has an effect at the lexical level (i.e., lexicality) when stimuli in the task are degraded. Despite other explanations are perhaps possible, we suggest that the whole pattern of data obtained when a factor influencing the recognizability of letters (i.e., SQ and TLC) is manipulated together with a lexical variable (i.e., word frequency and lexicality) can be explained by a dual-route emphasis account of list composition effects in reading.

7 LIST COMPOSITION EFFECTS IN DEGRADED READING

In the present chapter the effects due to list composition in degraded reading will be analyzed. The effects due to the type of stimuli presented in the task have been largely studied in reading aloud researches. Nevertheless, to the best of our knowledge, any published study has directly analyzed this issue when stimuli are degraded in the task. The research presented here aims to compare the reading performance to degraded English (regular) words and to degraded nonwords when presented in separated pure lists with the performance to the same stimuli when they are mixed together in the reading task. Our data will be discussed within the accounts traditionally proposed to explain list composition effects in reading.

7.1 Introduction

Whether and how humans can exert strategic control in reading tasks is a matter of debate in visual word recognition researches. For fluent readers, reading appears to be an extremely automatized process. However, strategic processes engaged in reading performance have been largely documented in studies analyzing the effects due to the composition of the list of stimuli in the task (e.g., Chateau & Lupker, 2003; Coltheart & Rastle, 1994; Kang, Balota, & Yap, 2009; Kinoshita & Lupker, 2003, 2007; Kinoshita, Lupker, & Rastle, 2004; Lupker et al., 1997; Monsell, Patterson, Graham, Hughes, & Milroy, 1992; Rastle & Coltheart, 1999; Tabossi & Laghi, 1992; Zevin & Balota, 2000). As a consequence, at least some aspects of the reading process must be strategically controlled by skilled readers.

The issue about strategic control in reading has been usually studied in terms of dual-route frameworks (e.g., Coltheart, 1978; Patterson & Morton, 1985, Coltheart et al., 2001), typically the DRC model. According to dual-route theories there are two ways to produce a phonological code. One way involves assembling a pronunciation based on knowledge of spelling-to-sound mapping; this strategy can be successfully used whenever the spelling-to-sound mapping of the letter string in input follows the standard rule of the language (i.e., regular words) and it is necessary for nonword reading. The other route involves accessing a lexical representation and retrieving the associated phonological code; this way can only be used for reading letter strings that have a lexical representation (i.e., words) and must assume a dominant role when words are irregular. Regardless of the characteristics of the input, both the routes are assumed to work in parallel on each stimulus. List composition effects are explained within dual-route models of reading by assuming that readers can selectively emphasized or de-emphasize the output of the two routes as a function of the type of stimuli presented in the task.

An alternative account of list composition effects in reading has been however proposed (e.g., Chateau & Lupker, 2003; Kinoshita & Lupker, 2003; Lupker et al., 1997), namely the time-criterion. This account relies on the idea that readers do not always initiate articulation as soon as possible; instead, skilled readers would set a flexible time-criterion (or deadline) for when articulation should start. Importantly, the position in time at which the criterion is set would depend on the type of stimuli presented in the task, thus explaining the effects due to list composition obtained in reading aloud researches.

Despite several attempts to distinguish between the two frameworks exist, these investigations have not yielded conclusive results. In fact, an extremely complex pattern of data has been obtained in these studies. A resume of the empirical findings is clearly far from the goals of the present dissertation; however, what is critical in this context is that whereas some of these studies reported evidence in favor of the route emphasis framework (e.g., Coltheart & Rastle, 1994; Kang et al., 2009) the results obtained in some other studies have been interpreted within a time-criterion account (e.g., Chateau & Lupker, 2003; Kinoshita & Lupker, 2003; Lupker et al., 1997). As a consequence, there is a certain agreement in considering the two hypotheses as not mutually exclusive; in other words, it could be that changes in the relative emphasis of a particular reading pathway and changes in the placement of a time-criterion jointly influence pronunciation performance. Hence, the two theories are today considered valuable approaches in interpreting list composition effects in reading.

To date, even if list composition found a large interest in researches on visual word recognition, these studies have been generally restricted to the investigation of these effects in standard viewing conditions. To the best of our knowledge, in fact, any published study has directly analyzed list composition effects in degraded reading. As discussed in the previous chapter, however, previous experiments involving the manipulation of a lexical variable under both the clear and the degraded presentation showed a different pattern of results (i.e., interaction vs. additive effects) depending on the type of stimuli presented in the task (e.g., Besner & O’Malley, 2009; Besner et al., 2010; O’Malley & Besner, 2008). Moreover, in the previous chapter, we also suggested that these findings may be explained by a dual-route emphasis account of list composition effects in reading.

Clearly, a direct investigation of the effects due to list composition when the stimuli are degraded in the reading task is required. In particular, in order to clarify the previous issues, one need to analyze reading performance to degraded words when they are solely presented in the task by comparing performance to the same stimuli when they are randomly mixed with degraded nonwords. Degradation may in fact determine particular effects of list composition in reading; if this will be the case, then these effects should be taken into account in any experiment involving a reduction of stimulus quality and modifying the composition of the list of stimuli.

This was the goal of the experiment reported below. In particular, the present study aims to compare the responses to (regular) words and nonwords when presented in pure list with the responses to the same stimuli when presented mixed together, when all the stimuli are degraded in the task. Specifically, three different conditions have been compared in the experiment:

a. a condition in which only degraded words were presented in the reading task;

b. a condition in which only degraded nonwords were presented in the reading task;

c. a condition in which degraded words and degraded nonwords (i.e., the same stimuli used in condition a e b) were randomly mixed in the reading task.

The three conditions have been alternated between participants, so that each subject performed either the conditions *a* (words only) and *b* (nonwords only) or the condition *c* (words mixed with nonwords). As a consequence, each item has been read only once by each participant (i.e., either in the pure condition or in the mixed condition). Moreover, also the order of the conditions *a* and *b* has been alternated between participants performing the task with pure lists of stimuli.

7.2 Method

Participants. Twenty-four students at the Macquarie University who had English as their first language and normal or corrected-to-normal vision participated as volunteers.

Design. Lexicality (words vs. nonwords) has been manipulated within subjects, whereas list composition (Condition: pure vs. mixed lists) has been manipulated between subjects.

Material. A total of 160 stimuli were used (these items can be seen in the Appendix, section F). They consisted of 80 regular monosyllabic English words and 80 legal monosyllabic nonwords with five letters in length. The two groups were matched in terms of phonological onset. The nonwords have been derived by changing a letter of an English regular word maintaining its initial phoneme. The words and the basewords which the nonwords were derived from were balanced for frequency (20.3 vs. 21.4 occurrences per million, $t<1$), orthographic neighbourhood size (4.6 vs. 4.5, $t<1$) and neighbourhood frequency (183 vs. 127, $t<1$). Moreover, words and nonwords were balanced in terms of neighbourhood size (4.6 vs. 4.4, $t<1$), neighbourhood frequency (183 vs. 124, $t=1.4$, n.s.) and TLC[31] (2.4 vs. 2.4, $t=1.2$, n.s.). Pure and mixed conditions as well as the order of presentation of the two pure lists have been alternated between participants, so that 1/3 the subjects read the only word block as the first list and the only nonword block as the second list, another 1/3 read the only nonword block as the first list and the only word block as the second list and the remaining subjects performed the reading aloud task including both words and nonwords.

Apparatus. The experiment took place in a sound attenuated and dim lit room. Stimuli presentation and data recording were controlled by DMDX software and running on a personal computer. The display was synchronized with the screen refresh cycle. Stimuli were presented centrally in upper-case letters in the 18-point Courier New font. All the stimuli were displayed in grey (RGB values: 3,3,2) on a black background (RGB values: 0,0,0). Responses were collected via a microphone connected to a voice key assembly. Response latencies were timed from stimulus onset to voice key activation, which also terminated the display.

Procedure. Participants were tested individually and sat in front of a computer screen. They were informed about the type of stimuli presented in the task and instructed to read each letter string aloud as quickly as possible minimizing errors. Each trial began with a 500 ms presentation of a fixation point at the centre of the computer screen followed by a 200 ms presentation of a blank. Immediately after the stimulus appeared and remained on the screen until a response was registered by the voice key or 3 sec elapsed. Stimuli were presented in a different random order for each participant. In mixed block stimuli order was controlled so to have the same number of words and nonwords within each group of ten stimuli. Responses were coded offline as correct or incorrect by the experimenter using CheckVocal software; the experimenter determined RTs using the waveform recorded by this software in order to reduce errors associated with voice key timing and correct for mistrial (i.e., voice key failure).

[31] LC was determined from empirical letter-confusion matrices obtained in previous studies (Gilmore et al., 1979; Loomis, 1982; Townsend, 1971; Van Der Heijden et al., 1984).

7.3 Results

An analysis based on linear mixed effects modelling (see Baayen, 2008; Baayen, Davidson, & Bates, 2008) has been performed in this study.

Mixed models extend the idea of traditional linear regression analysis that attempts to find out whether the distribution of a certain variable (response or dependent variable) can be, to a certain extent, predicted by a combination of others variables (explanatory or independent variables), and how the latter ones are affecting the former. The relationship between variables is modelled by fitting a linear equation to observed data on the assumption that the dependent variable is given by the weighted sum of the explanatory variables, plus some random noise. In the classical linear regression analysis only factors representing the so called fixed effects are incorporate in the model. Fixed effects are repeatable factors, which means that the set of possible levels for that factor is fixed and that each of these levels can be repeated. Usually, fixed factors correspond to the variables that are directly manipulated in an experiment. However, items and subjects are normally not repeatable. Items and subjects are sampled randomly from the population of stimuli and participants and replicating the experiment would usually involve selecting other items and other participants. This type of factors is called random effects because their levels are randomly sampled from a much larger population. Mixed models are models which incorporate both fixed and random effects. Random effects are assumed to be normally distributed with unknown variance, which will be estimated from the data; as Baayen (2008) states "*While fixed effects factors are modeled by means of contrasts, random effects are modeled as random variables with a mean of zero and unknown variance (…) In mixed models, the standard deviation associated with random effects are parameters that are estimated, just as the coefficients for the fixed effects are parameter that are estimated*" (p. 264).

Mixed modelling is particularly useful in psycholinguistic experiments in that allows to fit a linear equation to observed data by estimating a model in which the random effects for subjects and items are jointly analyzed. Hence, mixed models may offer a more powerful statistical analysis than traditional ANOVAs and be extremely useful especially when the variables (e.g., the factor representing list composition – condition: pure vs. mixed lists – in the present experiment) need to be manipulated between subjects. Another advantage of mixed modelling as compared to the traditional ANOVA is that other variables (i.e., factors that are not directly manipulated by the experimenter) can be added in the model in order to maximize the variance explained.

Linear mixed effects models are implemented with lme4 (Bates, Maechler, & Dai, 2008) and languageR (Baayen, 2008) packages in R development core team (2007).

Three different analyses have been performed both on correct reaction times and on errors data. A further analysis has been directed to analyze a restricted set of errors, i.e. lexicalization.

7.3.1 Analysis on list composition in function of lexicality

The first analysis was directed to analyze the effect of the variable Condition (pure vs. mixed lists) for the words and nonwords in the experiment. Mean RTs according to conditions and percentages of error are reported in Table 9.

	Nonwords		**Words**		Diff.	
Condition	RTs	%E	RTs	%E	RTs	%E
Pure	812	28.1	720	5.6	92	22.5
Mixed	748	18.3	671	3.6	77	14.7
Diff.	64	9.8	49	2		

Table 9. Mean reaction times (RTs) and percentages of error (%E) according to conditions.

RTs. First, we define a model (*m1*) with participants and items as random factors. Then, we define a second model (*m2*) by adding to model *m1* the factor trial-number (i.e., a factor representing the order in which the items have been presented in the experiment) as fixed factor. A formal comparison of *m1* and *m2* (namely, a log-likelihood test) showed a significant improvement in the model's fit, Chi2(1) = 9.9, $p < .001$. Since items have been presented in a different random order to each subject, the order of presentation may have affected each subject differently; we thus decide to allow the slope of the effect of trial to vary across subjects. To this purpose we centred the data and created a third model (*m3*) on them by adding to *m2* the by-subject random slope for trial (the correlation parameter hasn't been added[32]). A formal comparison between *m2* and *m3* showed a significant improvement in the model's fit, Chi2(1) = 128.9, $p < .001$. In a subsequent model (*m4*) we added Lexicality as fixed factor. A formal comparison of *m3* and *m4* showed a significant improvement in the model's fit, Chi2(1) = 53.7, $p < .001$. Then, we updated *m4* by adding Condition as a fixed factor but the increase in the model's fit was not significant, Chi2(1) = 1.4, p >

[32] A separated formal comparison between the models with and without the correlation parameter indicated that the model's fit doesn't improve when the correlation parameter is added, which suggests that it is not necessary in the model (see Baayen, 2008, p. 276).

.24. However, a further model (*m6*) has been tested by adding to the model *m4* the interaction between the two fixed factors (i.e., Lexicality and Condition). A formal comparison of *m4* and *m6* showed a significant improvement in the model's fit, Chi2(2) = 8.9, p = .01.

A further analysis was performed on *m6* to test the fixed factors effects. The effect of Lexicality was significant $|t| = 6.1$, pMCMC < .001, with words read faster than nonwords. Critically, the effect of Condition proved significant, $|t| = 1.5$, pMCMC ≈ .05, with items in the mixed list read faster than items in the pure list. Furthermore, the two factors were qualified by a significant interaction, $|t| = 2.8$, pMCMC < .001, with the effect of Condition larger for nonwords than for words.

Accuracy. First, we define a model (*a1*) with participants and items as random factors. Then, we define a second model (*a2*) by adding to model *a1* the factor trial-number as fixed factor but the model's fit was not significantly improved, Chi2(1) = 1.16, p > .28. However, a third model (*a3*) has been created by centring the data and adding to *a1* the by-subject random slope for trial (the correlation parameter has not been added). A formal comparison between *a1* and *a3* showed a significant improvement in the model's fit, Chi2(2) = 49.3, p < .001. In a subsequent model (*a4*) we added Lexicality as fixed factor. A formal comparison of *a3* and *a4* showed a significant improvement in the model's fit, Chi2(1) = 42.6, p < .001. Then, we updated the model *a4* by adding Condition as fixed factor and the increase in the model's fit was marginally significant, Chi2(1) = 3.4, p > .066. A further model (*a6*) has been tested by adding to model *a4* the interaction between the two fixed factors (i.e., Lexicality and Condition). A formal comparison of *a5* and *a6* showed an improvement in the model's fit, Chi2(1) = 7.1, p < .01.

A further analysis was performed on *a6* to test the fixed factors effects. The effect of Lexicality was significant $|t| = 5.3$, pMCMC < .001, with words more accurate than nonwords. Critically, the effect of Condition proved significant, $|t| = 2.7$, pMCMC < .01, with items in the mixed list read more accurately than items in the pure list. The two factors were qualified by a significant interaction, $|t| = 2.7$, pMCMC < .01, with the effect of Condition larger for nonwords than for words.

To conclude, the analysis on RTs showed a main effect of Lexicality (i.e., words are read faster than nonwords), a main effect of Condition (i.e., stimuli in the mixed list are read faster than stimuli in the pure list) and a significant interaction between the two factors (i.e., the effect of Condition is larger for nonwords than for words). The analysis on error rates is consistent with the RTs analysis, showing more accurate responses for the stimuli which are read faster.

The interaction between Condition and Lexicality that has been obtained indicates that the effect of Condition is larger for nonwords than for words. Since the main effect of Condition is significant, it follows that the difference between the pure and the mixed lists must be significant for nonwords. Therefore there is no need to test the effect of Condition for the nonwords by themselves[33]. However, one can't conclude from these results whether or not there is a significant effect of Condition for words. Hence, the effect of Condition just for words has been tested. We first define a model restricted to words only with participants and items as random factors (*w1*). Then, we add to model *w1* trial-number as fixed factor, but a formal comparison of *w1* and w*2* showed that the model's fit didn't significantly improve, Chi2(1) = 1.6, p >.21. In a third model we updated *w2* by centring the data and adding the by-subject random slope for trial (without adding the correlation parameter). A formal comparison between *w2* and *w3* showed a significant improvement in the model's fit, Chi2(2) = 6, p = .05. In a subsequent model (*w4*) we added Condition as fixed factor but the increase in the model's fit was not significant, Chi2(1) = .64, p = .42, thus suggesting that the effect of Condition is not significant for words.

7.3.2 Analysis on the order of presentation in pure lists

A separated analysis has been performed to determine whether the order in which the pure lists have been presented (words first vs. nonwords first) made any difference. For this purpose a subset of data corresponding to the RTs in the pure condition only has been extracted. Mean RTs and percentages of error are reported in Table 10.

	Nonwords		**Words**		Diff.	
Order	RTs	%E	RTs	%E	RTs	%E
First presented list	826	27	736	5.6	90	21.4
Second presented list	797	29.2	703	5.6	94	23.6
Diff.	29	-2.2	33	0		

Table 10. Mean reaction times (RTs) and percentages of error (%E) for Order and Lexicality.

[33] However, when this issue is directly investigated, the analysis shows that the effect of Condition is not significant for nonwords. In the analysis we first define a model restricted to nonwords only with participants and items as random factors (*nw1*); then, we add to model *nw1* trial-number as fixed factor. However, a formal comparison of *nw1* and *nw2* showed that the model's fit didn't significantly improve, Chi2(1) = .62, p >.43. In a third model we updated *nw2* by adding the by-subject random slope for trial (a formal comparison indicated that the correlation parameter is necessary) and a formal comparison between *nw2* and *nw3* showed a significant improvement in the model's fit, Chi2(3) = 8.2, p < .05. Finally we add, in a subsequent model (*mw4*), Condition as fixed factor but the increase in the model's fit was not significant, Chi2(1) = 1.5, p = .21, thus suggesting that the factor Condition has no significant effects for nonwords.

RTs. Following a similar procedure to the one explained in the previous analysis we first define a model (*o1*) with participants and items as random factors. Then, we add to model *o1* trial-number as fixed factor. The formal comparison of *o1* and *o2* showed a significant improvement in the model's fit, Chi2(1) = 17, p < .001. In a third model we updated *o2* by centring the data and adding the by-subject random slope for trial (the correlation parameter has not been added). A formal comparison between *o2* and *o3* showed a significant improvement in the model's fit, Chi2(1) = 110.1, p < .001. In a subsequent model (*o4*) we added Lexicality as fixed factor. A formal comparison of *o3* and *o4* showed a significant improvement in the model's fit, Chi2(1) = 51.3, p < .001. Then, we define a fifth model (*o5*) by adding Order as a fixed factor. A formal comparison showed that the model's fit significantly improves, Chi2(1) = 4.6, p < .05. We finally add the interaction between Lexicality and Order but the increase in the model's fit was not significant, Chi2(1) = 0, p = 1, thus suggesting that the interaction has no significant effects. A further analysis was performed on *o5* to test the fixed factors effects. The effect of Lexicality was significant |t| = 7.4, pMCMC < .001, with words read faster than nonwords. Also, the effect of Order proved significant, |t| = 1.5, pMCMC < .05, with the second list read faster than the first list.

Accuracy. First, we define a model (*ao1*) with participants and items as random factors. Then, we define a second model (*ao2*) by adding to model *oa1* the factor trial-number as fixed factor but the model's fit was not significantly improved, Chi2(1) = 2, p > .15. However, a third model (*ao3*) has been created by centring the data and adding to *ao1* the by-subject random slope for trial (the correlation parameter has not been added). A formal comparison between *ao1* and *ao3* showed a significant improvement in the model's fit, Chi2(2) = 39.8, p < .001. In a subsequent model (*ao4*) we added Lexicality as fixed factor. A formal comparison of *ao3* and *ao4* showed a significant improvement in the model's fit, Chi2(1) = 45.3, p < .001. Then, we define a fifth model (*ao5*) by adding Order as fixed factor, but a formal comparison showed that the model's fit didn't significantly improve, Chi2(1) = .67, p > .41. We finally add the interaction between Lexicality and Order but the increase in the model's fit was not significant, Chi2(2) = .66, p = .71. This analysis thus suggests that neither the effect of Order, nor the interaction between Order and Lexicality have significant effects on error data.

In conclusion, the main effect of Order that emerges in the RTs analysis is not surprising: participants read faster the stimuli that are presented as a second list for effect of practice with the task. Since there is no interaction involving Order, the data can be collapsed across this factor and hence the factor Condition has just has two levels, pure and mixed.

7.3.3 Analysis on word frequency

A separated analysis has been performed in order to analyze the effect of word frequency.

First, a correlation between the RTs to words and their frequency have been calculated: critically, the correlation proved significant when words are presented in pure list, r (80) = -.282, $p < .05$, but not when they are presented in mixed list, r (80) = -.07, n.s.

In mixed models analysis a subset of data corresponding to the words only has been extracted from the data. Furthermore, word frequency has been analyzed by adding a variable containing the specific frequency values in the model.

RTs. As in the previous analysis, we first created an initial model (*f1*) with participants and items as random factors. Then, we add to model *f1* trial-number as fixed factor but a formal comparison of *f1* and *f2* showed that the model's fit doesn't improve significantly, Chi2(1) = 1.6, $p > .20$. However, a third model (*f3*) has been created by adding to *f1* the random by-subject slope for trial (without the correlation parameter) and centring the data. A formal comparison of *f1* and *f3* showed a significant improvement in the model's fit, Chi2(1) = 5.7, $p < .05$. In a subsequent model (*f4*) we added Frequency as fixed factor. A formal comparison of *f3* and *f4* showed a significant improvement in the model's fit, Chi2(1) = 5.5, $p < .05$. Then, Condition as fixed factor has been added but the model's fit was not significantly improved, Chi2(1) = .65, $p > .42$. Finally, we created another model (*f6*) by adding the interaction between Frequency and Condition. A formal comparison of *f4* and *f6* showed a significant improvement in the model's fit, Chi2(2) = 8.5, $p < .05$.

A further analysis performed on *f6* showed that the interaction between Frequency and Condition proved significant, $|t| = 7.4$, pMCMC ≈ .005, with the effect of Frequency significant in the pure blocks, $|t| = 2.8$, pMCMC < .005, but not in the mixed blocks, $t < 1$.

Accuracy. First, we define a model (*af1*) with participants and items as random factors. Then, we define a second model (*af2*) by adding to model *af1* the factor trial-number as fixed factor but the model's fit was not significantly improved, Chi2(1) = .24, $p > .62$. Furthermore, a third model (*af3*) has been created by centring the data and adding to *af1* the by-subject random slope for trial (the correlation parameter was not necessary) but a formal comparison between *af1* and *af3* didn't show a significant improvement in the model's fit, Chi2(2) = .24, $p > .88$. In a subsequent model (*af4*) we added Frequency as fixed factor. A formal comparison of *af1* and *af4* didn't show a significant improvement in the model's fit, Chi2(1) = 1.7, $p > .19$. Finally, we defined a fifth model (*af5*) by adding the interaction between Lexicality and Frequency but the increase in the model's fit was not significant, Chi2(3) = 4.5, p > .21. This analysis thus suggests that neither the effect of Frequency, nor the interaction between Frequency and Lexicality have significant effects on error rates.

Moreover, the joint effects of word frequency and the order in which the words have been presented in pure blocks (i.e. first list vs. second list) have been analyzed in order to define whether the effect of frequency was modulated by the words being the first vs. the second presented list. For this purpose a set of data corresponding to words in the pure condition only has been extracted.

RTs. An initial model (*fo1*) with participants and items as random factors has been created. Then, we create *fo2* by adding to model *fo1* trial-number as fixed factor. However, a formal comparison of *fo1* and *fo2* showed that the model's fit doesn't improve significantly, Chi2(1) = .13, p > .72. A third model (*fo3*) has been created by adding to *f1* the random by-subject slope for trial (without the correlation parameter) and centring the data. A formal comparison of *fo1* and *fo3* showed a significant improvement in the model's fit, Chi2(1) = 10.4, p < .005. We then created the model *fo4* by adding Frequency as fixed factor. A formal comparison of *fo3* and *fo4* showed a significant improvement in the model's fit, Chi2(1) = 7.8, p < .01. In model *fo5* the fixed factor Order has been added but the model's fit was not significantly improved, Chi2(1) = .21, p > .65. Finally, a model (*fo6*) has been created by adding to *fo4* the interaction between Frequency and Order, but a formal comparison of *fo4* and *fo6* showed that the model's fit was not significantly improved, Chi2(2) = 3.5, p > .17. This analysis suggests that neither the main effect of Order nor the interaction between Order and Frequency have significant effects in the analysis.

Accuracy. We define a model (*afo1*) with participants and items as random factors. Then, we define a second model (*afo2*) by adding to model *afo1* the factor trial-number as fixed factor. However, the model's fit was not significantly improved, Chi2(1) = 0, p = 1. Furthermore, a third model (*af3*) was created by adding to *af1* the by-subject random slope for trial and centring the data (the correlation parameter was not necessary) but a formal comparison between *af1* and *af3* didn't show a significant improvement in the model's fit, Chi2(2) = .04, p > .98. In a subsequent model (*af4*) we added Frequency as fixed factor but a formal comparison of *af1* and *af4* didn't show a significant improvement in the model's fit, Chi2(1) = 2.2, p > .14. Finally, we define a fifth model (*af5*) by adding the interaction between Frequency and Order but the increase in the model's fit was not significant, Chi2(3) = 2.1, p > .54. This analysis thus suggests that neither the effect of Frequency, nor the interaction between Frequency and Order have significant effects on error rates.

To conclude, the analysis on RTs shows a main effect of Frequency and a significant interaction between Condition and Frequency, indicating that whereas the effect of word frequency is significant in the pure blocks it is not in the mixed blocks. This datum is consistent with the results obtained in the correlation analysis. Regarding the joint effects of Frequency and Order, the analysis showed that there is not interaction between these two effects; we thus conclude that the effect of frequency has the same amplitude when words are presented as the first list and when they are presented as the second list in the pure condition.

7.3.4 Analysis on lexicalization errors

A total of 14.9% of the data has been coded as errors in the experiment. Of these errors, 38% were fluent nonword errors and 45% were fluent word errors; the remaining 17% were non-fluent mistakes. We argue that it might be interesting to perform a separate analysis on the fluent word errors (i.e., lexicalization errors), since they might give important information in the context of list composition effects in reading.

Percentages of lexicalizations according to conditions are reported in Table 11.

Condition	**Nonword**	**Word**	Diff.
Pure	13.6	2.2	11.4
Mixed	5.9	0.6	5.3
Diff.	7.7	1.6	

Table 11. Percentages of lexicalization errors according to conditions.

In the analysis, we first define a model (*lex1*) with participants and items as random factors. Then, we define a second model (*lex2*) by adding to model *lex1* the factor trial-number as fixed factor but the model’s fit was not significantly improved, Chi2(1) = 1.9, p > .16. However, a third model (*lex3*) has been created by centring the data and adding to *lex1* the by-subject random slope for trial (the correlation parameter has not been added). A formal comparison between *lex1* and *lex3* showed a significant improvement in the model's fit, Chi2(1) = 15.6, p < .001. In a subsequent model (*lex4*) we added Lexicality as fixed factor. A formal comparison of *lex3* and *lex4* showed a significant improvement in the model’s fit, Chi2(1) = 25.7, p < .001. Then, we updated *lex4* by adding Condition as fixed factor and the increase in the model’s fit was significant, Chi2(1) = 6.5, p < .05. A further model (*lex6*) has been tested by adding the interaction between the two fixed factors and a formal comparison showed an improvement in the model's fit, Chi2(1) = 5.5, $p < .05$. A further analysis was performed on *lex6* to test the fixed factors effects. The effect of Lexicality was significant, $|t| = 3.6$, pMCMC < .001, with words more accurate than nonwords. Critically, the effect of Condition proved significant, $|t| = 3.4$, pMCMC < .005, with items in the mixed list read more accurately than items in the pure list. Furthermore, the two factors were qualified by a significant interaction, $|t| = 2.4$, pMCMC < .05, with the effect of Condition larger for nonwords than for words.

To conclude, the same pattern obtained for the whole set of error rates is obtained when lexicalization errors are separately analyzed. Hence, stimuli are more accurate (and there are less lexicalizations) in the condition in which they are read faster.

7.4 Discussion

The results of the present experiment can be summarized as follow.

1. Degraded words are read faster than degraded nonwords both in mixed and in pure blocks.
2. When words and nonwords are degraded, stimuli are read faster when mixed together than when presented in a pure list.
3. The mixed-list advantage is larger for nonwords than for words.
4. Nonwords are read less accurately than words. Mixed list are more accurate than pure list. The effect of list composition is larger for nonwords than for words in the analysis of errors.
5. The word frequency correlation with RTs is significant when words are presented in pure lists but not when they are mixed with nonwords; in addition, the effect of word frequency is significant in pure but not in mixed lists. In other words, frequency exerts an effect only when words are solely presented in the task.

The probably most interesting result obtained in this experiment is that reading latencies to degraded stimuli are faster when words and nonwords are mixed together in the task that when they are presented in separate pure lists. However, an issue remains to be clarified. Whereas the mixed-block advantage is significant for nonwords, it didn't prove significant for words in the mixed models analysis. We argue, however, that this latter result might be due to a lack of statistical power. In fact, when the effect of condition is analyzed for the nonwords themselves, it didn't prove significant as well. Since this is inconsistent with the other results (i.e., a main effect of Condition and a significant interaction between Condition and Lexicality) that could be an issue of low statistical power. Hence one needs to be cautious in concluding that there is not an effect of Condition with words. More evidence in favor of a significant effect of list composition for words is obtained in the ANOVA for items that shows faster RTs in the mixed than in pure lists, $F(1,156) = 103.6$, $MSE = 2298$, $p < .001$, and that this effect is significant both for words, $t(79) = 7.6$, $p < .001$, and for nonwords, $t(77) = 7.1$, $p < .001$. Even if further investigation is certainly needed, one can conclude from these results that RTs to words are not certainly faster in the pure list. Hence, if anything, words are read faster when mixed with nonwords that when solely presented in the task.

In the following sections the pattern of results that has been obtained will be discussed in the context of the two theories traditionally proposed to explain list composition effects in reading, i.e. the route emphasis account and the time-criterion account; moreover, a third interpretation – the lexical checking – will be considered.

7.4.1 The route emphasis account

A way to interpret the results that have been obtained is by considering the effects due to word frequency. Our analysis reveals that the correlation of word frequency with RTs proved significant when words are presented in pure list but not when they are mixed with nonwords. In addition and consistently with this result, the frequency effect proved significant in pure block but not in mixed block when analyzed with mixed modelling. Putting together, these data suggest that a lexical variable such as word frequency exerts its effect in the pure but not in the mixed lists.

This conclusion has important implications in interpreting our results within a dual-route emphasis account. According to this theory, in fact, readers would selectively emphasize or de-emphasize the output of one of the two routes. The effect of frequency would be explained either assuming that the lexical route is weaker in the mixed list, or hypothesizing that the non-lexical route is stronger in the mixed list. However, if the lexical route would be weaker in the mixed list, a pure-block advantage for words should have been observed. Instead, if anything, words are read faster in the mixed than in the pure lists in our experiment. Consider now the second hypothesis, that is the non-lexical route being stronger in the mixed list. Since both the routes can be used to read regular words, the emphasis on the non-lexical route in the mixed list could help word reading. Moreover, the mixed-block advantage obtained for nonwords is easily explained within this account because nonwords would be read faster and more accurate when the non-lexical route is stronger.

Finally, the mixed-block advantage has been shown to be larger for nonwords than for words: this result is simply accounted for, since the non-lexical route is particularly important in nonword reading.

To summarize, all the results obtained in our experiment are explained by assuming that the non-lexical route becomes stronger in the mixed list. The pattern of data is thus accounted for by a dual-route model as the DRC through a plausible and data-driven explanation.

7.4.2 The time-criterion account

According to the time criterion account, the moment in time when participants release the response would be displaced as a function of the difficulty of the stimuli they are exposed to: early with easy items, late with difficult items. Moreover, when easy and difficult stimuli are mixed together in the task, the criterion would be set at a position that is intermediate to the position used for the fast and slow stimuli when presented in pure list. This means that the responses to slow stimuli would be faster in mixed block than in pure block whereas the responses to fast stimuli would be slower in mixed block than in pure block. In our experiment words are read faster than nonwords in pure block: this means that words can be considered the easy (fast) items and nonwords the difficult (slow) stimuli. Consistently with our results, the time-criterion account predicts faster RTs to nonwords when they are mixed with words than in pure block. In addition, since the criterion is set to a position appropriate for the entire block of stimuli and the main effect is thus an homogenization of the RTs, the larger lexicality effect in pure than in mixed blocks as well as the larger frequency effect in pure than in mixed blocks are consistent with this hypothesis. However, the results we obtained for words cannot be reconciled with this account: according to the time-criterion, in fact, fast stimuli (i.e., words) should be read faster in pure block than when mixed with slower stimuli (i.e. nonwords). This prediction is clearly inconsistent with our experimental data.

Extending the time-criterion account to dual-route frameworks, Kwantes and Marmurek (2007) proposed that a manipulation of the reading aloud criterion not only determines the beginning of articulation but also affects how much information from the lexicon is used to create pronunciation. Specifically, when the response criterion is set low and naming latencies are short, the General Activation Strategy (or GAS) would be used: all the lexical entries that are similar to the target string in both spelling and sound contribute to reading. In contrast, when the response criterion is set high, the Specific Activation Strategy (or SAS) would be used since activation becomes less diffuse and pronunciation will be driven largely by its matching representation in the orthographic and/or phonological lexicons. Following this interpretation, one might assume that the GAS is used in mixed block whereas the SAS is preferred in pure block, since the precise identification of the target lexical unit would be less useful when half the stimuli are nonwords. Hence, reading latencies will be faster under the GAS than under the SAS, thus producing faster RTs in mixed list. However, an opposite pattern of error rates[34] would also be expected for words since these stimuli should be more accurate when a single lexical unit is activated in the lexicon (i.e., pure list). Instead, this is not the case in our experiment. Moreover, since the response to nonwords should not depend on how much lexical information is used to generate pronunciation, the pattern of results for nonwords is not clearly predicted by this account.

7.4.3 The lexical checking account

Another account, the lexical checking, may be relevant in this context. This theory has been proposed by Lupker et al. (1997; see also Kinoshita & Lupker, 2007 and Kinoshita et al., 2004) to explain the mixed block advantage they found for low-frequency irregular words and the pure-block advantage obtained for high-frequency irregular words when both the types of word are presented in pure block or mixed with nonwords. As Lupker et al. (1997) comment, "*the idea is that as the articulatory code builds up, readers can choose to consult an output lexicon to determine whether the phonological-code-generation process matches a code in their lexicon*" (p. 584). Of course, this strategy is more useful when frequency is low. According to the authors, this strategy would be invoked in pure blocks since participants must be sure that the articulatory code they produce is a word code; on the contrary, it is invoked less often in mixed blocks because half the stimuli are nonwords.

Even if regular words have been used in our experiment, we argue that a similar strategy could be plausibly used when the stimulus quality is reduced, since the letters that compose the string are not easily identified when degraded. When only words are present, readers could thus choose to consul the phonological output in order to determine whether it matches a code in their lexicon. On the contrary, this strategy would not be helpful in mixed blocks because half the stimuli are nonwords. To date, this account is hence very similar to the one previously discussed. Since a

[34] Error rates distribution is relevant also regarding the time criterion-account as originally proposed. In particular, an additional effect of naming slow stimuli more rapidly and naming fast stimuli more slowly should be reciprocal changes in error rates (Strayer & Kramer, 1994): when slow stimuli are named more rapidly, numerically larger error rates should be observed, whereas naming fast stimuli more slowly should lead neither a decrease nor an improvement in accuracy. However, our data showed an opposite pattern. In fact, in our experiment, the slowest stimuli (i.e., nonwords) are clearly more accurate in the condition in which they are read faster (i.e., mixed block).

further operation is needed when words are solely presented, RTs would be slower. This hypothesis is thus consistent with RTs to words faster in mixed list. Furthermore, since this strategy would be particularly useful when frequency is low, one might expect the mixed-block advantage being larger for low-frequency words than for high-frequency words, a prediction that match our empirical data.

However, this account doesn't predict anything about nonword data, thus being unable to explain the mixed-block advantage we obtained for these stimuli. Also, the error rates distribution is inconsistent with this hypothesis. According to the lexical checking account, in fact, a numerically larger error rate is expected for words in mixed lists. Since readers would not consult an output lexicon in mixed block, errors in word reading should be more frequent than in pure blocks. Critically, the data obtained in our experiment didn't meet this prediction.

7.4.4 Simulation

In the previous sections we argued that the results obtained in our experiment might be explained thought the route emphasis account by assuming that the non-lexical route becomes stronger in mixed lists. In this section DRC model simulations directed to test this hypothesis will be present.

The items used in the experiment were run through the DRC 1.2 under the degraded condition. To simulate the reduction in stimulus contrast the strength of the connections between features and letters has been reduced by 40%. Specifically, the feature-to-letter excitation was reduced to .003, whereas the feature-to-letter inhibition was reduced to .09.

First, a simulation in which only a reduction in stimulus quality was implemented has been performed. This data would simulate the pure condition of our experiment. The model did not produce any errors. Mean cycles to criterion are reported in Table 12 (Pure Condition).

Then a series of simulations have been performed in order to determine how the strength of the non-lexical route could be manipulated to simulate the mixed condition. In the DRC 1.2 the operations of the non-lexical route are regulated by three parameters, namely the GPC-Phoneme-Excitation, the GPC-Critical-Phonology and the GPC-Onset.

The GPC-Phoneme-Excitation determines the amount of activation that reaches the phonemic buffer from the GPC system: higher levels of activation determine a stronger influence of the non-lexical route on pronunciation. As a consequence, the higher this value is, the major the strength of the non-lexical route.

The GPC-Critical-Phonology controls the level of activation that has to be reached in each phoneme unit before the GPC route moves on the next letter of the string. In other words, this parameter regulates the speed in which the non-lexical route moves from left to right along the string of letters; the higher this value is, the weaker the non-lexical route[35].

Finally, the GPC-Onset determines how many cycles after the perceptual processing (i.e., feature and letter levels analysis) the non-lexical route starts to operate; the higher this value is, the weaker the non-lexical route.

Our simulations suggested that the mixed-block advantage obtained in our experiment can in fact be simulated by the DRC model by increasing the strength of the non-lexical route. In particular, mean cycles to words and nonwords are progressively lower as either the parameter GPC-Phoneme-Excitation is increased or the parameter GPC-Critical-Phonology is reduced. However, the parameter that seems to have a major role in determining this effect is the GPC-Phoneme-Excitation, rather than the GPC-Critical-Phonology. As a consequence, we will present the data we obtained by manipulating the former parameter. In particular, in our simulation the GPC-Phoneme-Excitation has been increased from its default value of .051 to .1 in order to simulate the mixed condition of our experiment. The model did not produce any errors. Mean cycles to criterion are reported in Table 12 (Mixed Condition).

	Nonwords	**Words**	Diff.
Condition	Cycles	Cycles	Cycles
Pure	166.5	78.7	87.8
Mixed	114.2	70	44.2
Diff.	52.3	8.7	

Table 12. Mean cycles for Condition and Lexicality.

An ANOVA with Condition as repeated factor and Lexicality as independent factor was conducted on cycles. The DRC behaviour mimed that of humans. Analysis showed a main effect of Condition, $F(1, 158) = 22831$, $MSE = 3.3$, $p < .001$, a main effect of Lexicality, $F(1, 158) = 26567$, $MSE = 13$, $p < .001$, and a significant interaction, $F(1, 158) = 11622$, $MSE = 3.3$, $p < .001$, imputable to the size of the effect of Condition being larger for nonwords than for words. Paired comparisons reveal that the effect of Condition (pure vs. mixed lists) is significant both for words, $t(79) = 61.7$, $p < .001$, and for nonwords, $t(79) = 138.2$, $p < .001$. Also, the effect of Lexicality is significant both in the pure block, $t(158) = 161.6$, $p < .001$, and in the mixed block, $t(158) = 130.7$, $p < .001$.

[35] This parameter has been introduced in the newest version of the DRC model and substitutes the parameter called GPC-Interletter-Interval assumed in the version of the model originally presented (the DRC-PR) which also controlled the left-to-right movement of the non-lexical route; specifically, the GPC-Interletter-Interval controlled after how many processing cycles the GPC system moves on the next letter in the string.

A separate analysis has been performed in order to analyze the effect of word frequency. In this analysis the words have been median-split on frequency to create two groups balanced on all the other relevant psycholinguistic variables[36].

Mean cycles to criterion are reported in Table 13.

	Low-Frequency	**High-Frequency**	Diff.
Condition	Cycles	Cycles	Cycles
Pure	80.6	76.9	3.7
Mixed	71.4	68.6	2.8
Diff.	9.2	8.3	

Table 13. Mean cycles for Condition and Frequency.

An ANOVA with Condition as repeated factor and Frequency as independent factor has been performed. The DRC 1.2 behaviour mimed that of humans. Analysis showed a main effect of Condition, $F(1, 78) = 4385$, $MSE = .698$, $p < .001$, a main effect of Frequency, $F(1, 78) = 117.8$, $MSE = 3.5$, $p < .001$, and a significant interaction, $F(1, 78) = 12.9$, $MSE = .698$, $p = .001$, imputable to the size of the word frequency effect being smaller in the mixed than in the pure blocks[37].

Summarizing, when stimulus contrast is manipulated by reducing the strength of the connections between the feature and the letter units and the mixed condition is simulated by increasing the GPC-Phoneme-Excitation parameter, the following results are obtained in the simulation of the DRC 1.2 model:

1. words are read faster than nonwords, both in the pure and in the mixed blocks;
2. words are read faster in mixed than in pure blocks;
3. nonwords are read faster in mixed than in pure blocks;
4. the mixed-list advantage is larger for nonwords than for words;

[36] The 80 words used in the experiment have been divided into two groups of 40 stimuli balanced for orthographic neighbourhood size (4.5 vs. 4.8, $t<1$) and neighbourhood frequency (141 vs. 226, $t=1.16$, n.s.), one containing low-frequency words (mean: 2.8; range: 0.6-5.8) and the other including high-frequency words (mean: 37.8; range: 6-112).

[37] Differently from human performance, the effect of frequency is significant both in pure block, $t(78) = 10.2$, $p < .001$, and in mixed block, $t(78) = 9.7$, $p < .001$, in the DRC model's simulation. However, consistently with our experimental data, it is easily demonstrable that both the frequency and the lexicality effects are progressively reduced as the non-lexical route becomes more and more strong by increasing the value of the GPC-Phoneme-Excitation parameter. We argue this is an important argument in favour of our interpretation.

5. the frequency effect is larger in pure than in mixed blocks.

In conclusion, all the results obtained in our experiment are correctly simulated by the DRC model by increasing the strength of the non-lexical route in mixed condition.

7.4.5 Conclusion

Our experiment showed that in degraded condition reading latencies are faster when words and nonwords are mixed together than when they are presented in pure lists. Whereas this effect is significant for nonwords, the mixed-block advantage for words needs to be further investigated. We argue that the effects of list composition obtained in degraded nonword reading can be explained through a route emphasis account by assuming that the non-lexical route becomes stronger in mixed lists. Consistently with this interpretation, DRC model's simulations have been shown to reproduce the whole pattern of data when the non-lexical route is made stronger by increasing the parameter regulating the GPC-Phoneme-Excitation.

Moreover, following the rationale of the present thesis, the results obtained in this experiment have important implications for the account proposed in the previous chapter. In fact, in Chapter 6, we suggested that the whole pattern of data obtained in reading experiments manipulating a factor influencing the recognizability of letters (i.e., SQ and TLC) together with a lexical variable (i.e., lexicality and word frequency) may be accounted for by the DRC model assuming that the balance between the two routes favours the lexical route over the non-lexical route more strongly when only words are presented in the task than when words are randomly mixed with nonwords. This balancing may in fact consists in the non-lexical route being stronger in mixed list, as suggested in the present experiment: when the non-lexical route is emphasized, in fact, the response mostly depends on this route and the effect of the lexical route will be weaker.

This account is also consistent with the general idea that readers might exert strategic control on the non-lexical route rather than on the lexical route (see, e.g., Herdman, 1992; Paap & Noel, 1991). In fact, whereas processing on the lexical route is based on direct and automatic association between orthographic forms and their phonologies, the non-lexical route involves the assembly of phonological segments and it presumably requires more resources. Hence, it would be the non-lexical route, rather than the lexical route, that is strategically emphasized or de-emphasized in the task. On the contrary, the lexical route would be highly automatized and less susceptible to strategic influences.

Regardless of the fact that an explanation in terms of dual-route framework is possible, the pattern of results obtained in this experiment remains somehow very puzzling. If we think that what subjects do is to have one set of parameters in the pure condition (call this set *X*) and a different set of parameters in the mixed condition (call this set *Y*), then our results suggested that set *X* is not optimal and set *Y* is better (both in speed and accuracy of performance). Thus, why didn't subjects use the set *Y* both in pure and in mixed conditions? In other words, if the balancing between the two routes used in mixed condition is the optimal strategy, why isn't it used also in the pure list? Of course, our study is just a first attempt to analyze list composition effects in degraded reading and this and other issues need to be further analyzed.

Despite further researches are certainly needed, the pattern of results we obtained is interesting and the explanations investigated are promising. New effects of list composition in reading have in fact been documented when stimuli are degraded in the task. Moreover, important implications for the studies manipulating stimulus quality together with another variable in function of the type of stimuli presented in the task have been pointed out. Future researches in this direction will provide further knowledge on visual word recognition and reading aloud when stimuli are presented in not-optimal visibility conditions.

8 GENERAL DISCUSSION

A number of highly successful computational models of reading implement multiple levels of representation that get activated when a letter string is presented. A central feature of these models is that activation is usually assumed to spread in a cascaded fashion across the different levels of processing. In systems that operate by cascaded processing, there are no thresholds within levels and, as soon as even a small amount of activation is accumulated in an early stage, this flows on to later levels.

Despite cascaded processing in visual word recognition is commonly accepted, recent experiments with skilled readers involving the manipulation of a factor affecting the rate of processing (i.e., stimulus quality) in conjunction with another variable produced results that are not easily reconciled with this account. Critically, cascaded processing has been evaluated in these studies by referring to the DRC model, perhaps the most successful computational model of reading aloud and visual word recognition.

One of the most relevant results in this context has been obtained by Besner and Roberts (2003). The authors showed that the letter length effect and the effect of SQ are additive in nonword reading tasks, so that the length effect has the same amplitude regardless of the level of SQ. On the contrary, an interaction of the two factors, with the length effect smaller for degraded than for clear stimuli, is obtained in the DRC model's simulation. According to the model, in fact, activation is continuously accumulated during phonemic processing and, since reading longer nonword requires more time, activation should grow more for longer letter strings; therefore, the delay produced by a degraded stimulus should be partially reduced for longer nonwords. To allow the DRC model to fit with the empirical results, Besner and Roberts (2003; see also Blais & Besner, 2007) have proposed to change the cascaded assumption so that activation would spread in a thresholded fashion for the non-lexical route. In particular, they proposed to threshold the output of the letter level, since "*If the letter-level module does not pass activation to the grapheme-phoneme conversion process until a threshold is reached, the interaction in the simulations (...) would likely not occur. Instead, additive effects of processing rate and letter length would be expected*" (Besner & Roberts, 2003, p. 403).

Moreover, Reynolds and Besner (2004) analyzed the joint effects of SQ and orthographic neighbourhood size (N) on nonword reading in both skilled readers and in the DRC computational model. The authors reported that, while N and SQ exert additive effects on skilled readers' latencies, these factors interact in the DRC model simulations, with the effect of SQ being smaller for high-N nonwords than for low-N nonwords. This interaction is caused by the interactive activation between the letter level and the orthographic lexicon assumed in the model: since processing is cascaded, the lexical entries corresponding to the orthographic neighbours of the nonword in input are activated and in turn feed activation back to the letter level; the effect of degradation would be thus reduced as the number of orthographic neighbours of a nonword increases. The results reported by Reynolds and Besner (2004) appear therefore to be inconsistent with cascaded processing and a threshold at the letter level has been proposed; differently from the previous hypothesis, however, also the lexical route would be activated by the output of the letter level analysis in a thresholded fashion.

Inconsistent evidence with the DRC model's cascaded assumption has been obtained also in word reading. O'Malley and Besner (2008) showed that when only words were presented in the task, the effects of SQ and word frequency interacted, with the effect of SQ larger for low-frequency words than for high-frequency words; this result is consistent with a cascaded account. However, when also nonwords were included within the experimental stimuli, word frequency and SQ exerted additive effects on reading latencies. This latter result is inconsistent with models operating in a purely cascaded fashion. In addition, SQ and lexicality have been shown to exert additive effects on skilled readers' latencies (Besner & O'Malley, 2009; Besner et al., 2010; O'Malley & Besner, 2008), further contrasting cascaded processing. The authors referred again to a threshold at the letter level in order to explain their results. Specifically, it has been proposed that readers might switch from cascaded processing to thresholded processing as a function of the experimental context (i.e., lexicalization hypothesis). In particular, when stimuli are degraded and words and nonwords are mixed together in the task, participants would threshold the output of the letter level to prevent lexical capture of nonwords; instead, when the experimental list is composed of words only, processing would flow in a purely cascaded fashion. Lexical capture might in fact occur in the DRC model when stimuli are degraded since a nonword may activate a lexical entry enough to be erroneously read as the word instead of the nonword. A threshold at the letter level should prevent lexical captures because activation would pass to later stages only once the letters comprising the stimulus have been uniquely identified: thus the likelihood of erroneously selecting a word given a nonword is reduced or eliminated.

To date, when the letter level is implemented as thresholded in the DRC model, the additive effects of SQ and letter string length and of SQ and N in nonword reading as well as the additive effects of SQ and word frequency would be in fact correctly simulated by the model (see Besner et al., 2003), thus proving that this modification is successful.

We argue that hypothesizing thresholded processing at the letter level is not an account without issues. Empirical data inconsistent with this proposal have been in fact documented; for example, in order to explain the additive effects of SQ and N, one need to assume that the model is thresholded before it activates the lexical route; this account is clearly inconsistent with empirical data indicating that SQ and repetition interact (Blais & Besner, 2007) as well as SQ and semantic priming interact (e.g., Ferguson et al., 2009) when reading words. More generally, it has not been yet demonstrated whether implementing the letter level as thresholded would allow the DRC model to simulate all the effects that its current computational version does simulate: in science, any new theory should instead be demonstrated able to account for all the results that a previous theory (or a previous version of the same theory) was able to explain, plus some other empirical data. Furthermore, a whole change of the DRC model's cascaded nature is intrinsic in this proposal. This solution assumes, in fact, that at least some levels in the reading system are discrete and serially organized and that information processing is at least partially thresholded. As a consequence, accepting this modification would mean to reject the solution of cascaded processing today widespread accepted in researches on visual word recognition.

8.1 Review of the main findings

The main goal of the researches presented in this thesis has been to test the hypothesis of a threshold in the reading system. Following the outline of the final section of Chapter 1, I will summarize the main results that have been presented in the previous chapters. I will start by summarizing the findings obtained in nonword reading tasks (Chapter 2 to 5); then, I will focus on the results obtained in reading tasks manipulating the composition of the list of stimuli (Chapter 6 and 7).

8.1.1 Factorial manipulations in nonword reading

The joint effects of letter string length and orthographic neighbourhood size (N) have been analyzed in the first experiment reported in this thesis (Chapter 2) in clear (i.e., non-degraded) presentation by employing an Italian nonword reading task. The relevance of this experiment for the purposes of the present thesis is evident in that both letter string length and N have been shown to exert additive effects with SQ in nonword reading (see Besner & Roberts, 2003, and Reynolds & Besner, 2004, respectively); critically, these results have been interpreted as evidence against cascaded processing in the reading system. In our experiment, an interaction between letter string length and orthographic neighbourhood size (N) has been obtained, thus supporting a cascaded model with a dual route architecture like the DRC. These results are in fact incompatible with the postulation of a threshold in the reading system: the interaction between length and N in nonword reading can be explained only assuming that the (parallel) lexical route determines, together with the (serial) non-lexical route, the pronunciation of the nonword in input; crucially, lexical influence in nonword reading is a strong evidence in favour of cascaded processing.

In Chapter 3, the effect of a variable already noted in researches on letter-by-letter dyslexia, the Total Letter Confusability (TLC), has been analyzed following the assumption that, when stimuli are degraded, the letters are difficult to identify and their visual similarity could hence become important. The role of the TLC has been assessed for Italian skilled readers when reading clear and degraded nonwords. The results obtained in our experiment indicated that TLC influences healthy readers' performance when stimuli are degraded. This finding has relevant implications for reading researches analyzing the effect of psycholinguistic variables when jointly manipulated with stimulus quality. Since high-LC letters suffer more when degraded than low-LC letters, TLC is an important factor to consider in experiments involving degraded presentation of letter strings, a possibility which the authors did not consider in previous studies.

Following the result described above, the experiment reported in Chapter 4 was directed to analyze whether the additive effects of SQ and letter string length reported by Besner and Roberts (2003) in nonword reading may depend on a confounding between TLC and letter string length. In the experiment reported by Besner and Roberts (2003), in fact, TLC was not controlled; since TLC typically increases as letter length increases, it follows that the long letter strings used in this study had higher confusability values than the short letter strings. Moreover, the data we obtained in the experiment demonstrated that this confounding is in fact responsible of the additive effects that have been previously observed. More specifically, when the short and the long nonwords used in the task are matched for TLC, the effect of SQ results smaller for longer nonwords, as predicted by cascaded processing; instead, when the TLC is left uncontrolled, the opposite pattern of results has been obtained. In other words, the interaction between SQ and letter string length that is simulated by the DRC model is in fact obtained for human readers under the adequate experimental conditions (i.e., when short and long stimuli are matched in terms of TLC). Clearly, this finding strongly sustains cascaded processing in reading: a similar interaction would be in fact totally prevented by postulating a threshold at the letter level, thus contradicting our results.

In Chapter 5 we focused on the joint effects of SQ and orthographic neighborhood size (N) in nonword reading. Reynolds and Besner (2004) obtained that low-N and high-N nonwords were harmed by stimulus degradation to the same extent. In order to explain this additivity within the DRC model, a threshold at the letter level has been proposed; however, this account is clearly inconsistent with previous well-established data (see, Blais & Besner, 2007; Ferguson et al., 2009). In order to further test this hypothesis we performed an English nonword reading task by manipulating two factors: stimulus quality and whether nonwords have or not orthographic neighbors. The previous results may in fact be due to the particular manipulation of the variable N since both low-N and high-N nonwords would determine interactive activation between letter units and the lexicon that in turn would reduce the effect of degradation. Consistently with our prediction, an interaction between SQ and N (zero-N vs. many-N) has been obtained in our study. Clearly, this result is explained only by assuming cascaded interactive activation in the reading system. On the contrary, if there would be a threshold at the letter level, the presence/absence of orthographic neighbors should not play any role on stimulus degradation and the two effects should be thus additive, contrary to our empirical data. Moreover, an interpretation of the additivity previously obtained in terms of a confounding with TLC has been proposed in this chapter. This interpretation is partially supported by the analysis on the material used in the Reynolds and Besner’s (2004) experiment: since their high-N nonwords had higher confusability values than their low-N nonwords, the effect of degradation for high-N stimuli could be in fact smaller than how reported and the true result could be thus an interaction between the two variables. Clearly, further investigation is needed in order to define whether these results are in fact caused by TLC or rather depend on the different manipulation of the variable N.

In conclusion, all the results obtained in these experiments require activation to spread in a cascaded fashion in the reading system, thus denying the hypothesis of a threshold at the letter level. Moreover, we argue that the additivities of SQ and another factor obtained in previous nonword reading tasks would be due to a confounding with TLC and that when this confounding is removed the true results would be the interactions predicted by cascaded processing. Even if further work is needed in this context, the data obtained so far suggest this may be in fact the case.

8.1.2 Factorial manipulations in reading tasks as a function of list composition

The factorial manipulations of SQ and lexical variables in function of the presence/absence of nonwords in the task have been taken into account in Chapter 6. Previous studies showed that the joint effects of SQ and a lexical variable (e.g., word frequency) are modulated by the composition of the list of stimuli in the task: more specifically, the factors would interact when only words are presented but their effects would be additive when participants read both words and nonwords (e.g., O'Malley & Besner, 2008). The pattern of findings obtained in this context has been explained in previous literature by hypothesizing that when lexical capture may occur (i.e., the task comprises both words and nonwords and stimuli are degraded) the letter level is thresholded (i.e., lexicalization hypothesis). The main aim of our study was to test this hypothesis. We argue, in fact, that SQ may be not an adequate manipulation to test whether the letter level is thresholded since this variable affects the feature level analysis (i.e., not the letter level): instead, if one wants to test whether there is a threshold specifically at the letter level, then a variable directly affecting this level should be considered. To this purpose we manipulated Total Letter Confusability (i.e., a variable that affects the letter level) and lexicality within an Italian reading aloud task where all the stimuli were degraded. According to the lexicalization hypothesis the letter level should be thresholded in these experimental conditions and hence additive effects between the two factors are expected. Critically, TLC and lexicality instead interact in our experiment: whereas TLC affects degraded nonword reading, it does not affect degraded word reading. We hence concluded that the letter level is not always thresholded when the conditions postulated by the lexicalization hypothesis are met, thus confuting this account; on the contrary, our finding strongly supports cascaded processing in the reading system. A more general conclusion that emerges from this study is the following: TLC may be an useful second manipulation (alternative to SQ) in factorial experiments in that it allows to directly test whether processing is thresholded specifically at the letter level analysis, as it has been typically proposed.

Since TLC does not affect word reading, it follows that the additivities obtained in these studies cannot depend on this variable: a different interpretation of these results is hence needed. Specifically, we argue that these data may depend on the strategic control that readers exert in reading as a function of the type of stimuli presented in the task. This argument depends upon the claim that list composition has an effect in degraded reading, an hypothesis that has been tested in Chapter 7. In particular, in our last experiment, words and nonwords were presented either in pure lists or mixed together in an English reading aloud task; critically, all the stimuli were degraded. Using mixed modelling analysis, we showed that stimuli were read faster and more accurately when randomly mixed in the task than when presented in separated pure lists. These data have been interpreted through a dual-route emphasis account by suggesting that the non-lexical route becomes stronger in mixed condition: since words in our experiment were regular, in fact, the lexical route could help both word and nonword reading. Consistently with this interpretation, the simulation of these results with the DRC model mimed the human performance when the mixed condition is reproduced by strengthening the non-lexical route. Regardless of that, however, participants seem to behave in an extremely irrational way in this experiment; in particular, since the strategy used when words and nonwords are randomly mixed in the task is optimal, it is not clear why subjects did not do the same also when words and nonwords are presented in separated pure lists. In other words, even if our results can be interpreted within a theoretical framework, it is instead not clear why participants would perform the task in such a way: answering this question will be an interesting challenge for feature researches.

Nevertheless, following the previous reasoning, the results obtained in Chapter 7 prove the existence of particular effects of list composition when stimuli are degraded in the task. As a consequence, these effects have to be taken into account by experiments involving factorial manipulations of SQ which also varied the type of stimuli presented within the experimental list. In particular, we suggest that the additivities obtained when SQ is manipulated together with a lexical factor in tasks comprising both words and nonwords may be explained by referring to a route emphasis account of list composition effects in reading; also, the solution proposed to explain the mixed-block advantage obtained in Chapter 7 (i.e., the non-lexical route is emphasized when words and nonword are mixed in the task) is – at least from a theoretical point of view – adequate also to explain the whole pattern of data obtained in this kind of studies. Note that we are not necessarily suggesting that what participants did is exactly the same in these experiments: there is, in fact, a main difference between our task in which all the stimuli were degraded and previous studies manipulating SQ where half the stimuli were clear and the other half degraded in the task; rather, we argue that participants may strategically control the balance of the two functional routes in function of list composition in a particular way when (at least part of) the stimuli in the task are degraded.

In conclusion, we showed that postulating thresholded processing in the reading system is not an adequate solution in that this account is not reconciled with our empirical findings. In alternative, we suggest that the data obtained by factorially manipulating SQ and a lexical factor in function of list composition may depend on the strategic emphasis that readers give to the lexical and to the non-lexical routes while reading in response to the type of stimuli presented in the task. A study directed to analyze list composition effect in degraded reading partially supported this interpretation. Even if additional work is needed to further investigate this issue and to define whether the solution proposed is adequate to interpret the whole pattern of empirical findings, the data collected so far strongly sustain a dual-route framework assuming cascaded processing.

8.2 Interpretation within the Dual-Route Cascaded model

In this section, I will focus on the discussion of the results (either additivity or interaction) obtained when SQ is factorially manipulated with another variable by referring to the Dual-Route Cascaded model of reading. Table 14 provides a summary of the data at present published in this context as well as of the results presented in this thesis in function of the type of stimuli presented in the task.

Lexical / Non-lexical factor	**Stimulus quality**		
	Pure list (words only)	**Pure list (nonwords only)**	**Mixed list (words and nonwords)**
Letter length		*Additive effects* (Besner & Roberts, 2003) *Interaction* (Chapter 4) SQ effect larger for short than for longer items	
Neighbourhood density		*Additive effects* (Reynolds & Besner, 2004) *Interaction* (Chapter 5) SQ effect larger for zero-N than for many-N items	
Word frequency	*Interaction* SQ effect larger for low-frequency words (O'Malley & Besner, 2008)	–	*Additive effects* (O'Malley & Besner, 2008)
Lexicality	–	–	*Additive effects* (e.g., Besner et al. 2010)
Semantic priming	*Interaction* SQ effect larger for unrelated vs. related target (e.g., Ferguson et al., 2009)	–	
Repetition			*Interaction* for exception words *Additive effects* for nonwords (Blais & Besner, 2007)
Regularity	*Interaction* SQ effect larger for regular than for exception words (Besner et al., 2010)	–	*Additive effects* (Besner et al., 2010)

Table 14. Results of experiments involving the factorial manipulation of SQ and various lexical and non-lexical factors when reading aloud as a function of list composition.

We argue that all the data reported in the previous table are consistent with a cascaded framework and can be in particular interpreted within the DRC model.

Consider first the experiments analyzing the joint effects of SQ and another variable (letter string length and N) in nonword reading. As said, our assumption is that the additivities previously observed (Besner & Roberts, 2003; Reynolds & Besner, 2004) are due to a confounding with the TLC, a variable that is involved in reading when stimuli are degraded (see Chapter 3); we also hypothesized that when this confounding is removed the results are consistent with the DRC model's predictions. Even if further empirical work is needed, the results obtained so far strongly corroborate this interpretation. In particular, the additive effects of SQ and letter string length in nonwords reading depend indeed on a confounding with TLC and when this confounding is removed the two factors interact with the effect of SQ being smaller for longer nonwords (Chapter 4), a result that is clearly predicted by the DRC model. Moreover, we suggest that a similar confounding could also explain the additivity of SQ and N that has been documented; whether this hypothesis is valid remains to be demonstrate, even if the analysis on the stimuli that have been used provides partial support to this interpretation. Importantly, we showed that SQ and N interact when nonwords with and without orthographic neighbours balanced in terms of TLC are presented in the task (Chapter 5), perfectly matching the DRC model's interpretation.

Consider now the pattern of effects that SQ has with lexical variables in function of the presence/absence of nonwords in the task. In this dissertation we directly focused on the joint effects of SQ and word frequency/lexicality (Chapter 6). In general, we argue that these data can be explained within the DRC model by referring to a route emphasis account of list composition effects in reading. Following the results obtained in Chapter 7, we specifically proposed that the non-lexical route may be emphasized in mixed list compared to when words are solely presented in the task. As a consequence, the lexical route would have a weaker effect on pronunciation when nonwords are present than when they are not.

The interpretation of the previous results within the DRC framework would be as follow.
Since degradation affects the feature level analysis, activation from the orthographic lexicon acts indirectly on this variable: in fact, since there are no connections in the model from the letter to the feature units, the feedback from the lexicon can have an effect on SQ only acting at the level of letter identification. When only words are presented in the task, the feedback from the lexicon may have an effect on the letter level relatively early during the process that allows pronunciation, thus reducing the effect of degradation. In other words, since the lexical route is relatively fast as compared to the non-lexical route, the former routine may have an important role in demining pronunciation performance: hence, an interaction is expected between SQ and a lexical variable

such as word frequency when only words are present. On the contrary, when words and nonwords are mixed together, the non-lexical route would be strengthened. It follows that responses mostly depend on this route and the feedback from the lexicon may thus have a later effect on pronunciation, thus being unable to act on degradation; SQ and word frequency would be thus additive when words and nonwords are randomly mixed in the task; moreover, the effect of SQ will be additive also with lexicality. Further support to this interpretation is obtained by considering the simulations reported in Chapter 7: here the effects of word frequency and lexicality were in fact reduced by strengthening the non-lexical route.

In Chapter 6 we also demonstrated that TLC and lexicality interact when words and nonwords are randomly mixed in the task and all the stimuli are degraded. We argue that also this effect is consistent with the previous hypothesis. In particular, this interaction would be explained because TLC affects a different level of processing: whereas the effect of SQ is at the feature level, TLC has its effect at the subsequent level of letter identification. As a consequence, the feedback from the lexicon acts directly on TLC and its effect could arise early enough in pronunciation also when words and nonwords are mixed in the task: hence TLC and lexicality interact.

As reported in Table 14 at least three other results are relevant in this context. One of them is the interaction between SQ and semantic priming in word reading: as said, this interaction is perfectly predicted by the DRC model and it is consistent with its cascaded assumption; hence, this result does not require to be further discussed. However, other two results (i.e., the joint effects of repetition, SQ and lexicality and the joint effects of SQ and regularity) need additional consideration. For the sake of completeness, these experiments will be discussed in details in the next section.

Moreover, even if the previous findings could all be explained within a cascaded framework, the DRC model is not yet been demonstrated able to simulate the whole pattern of data. As discussed in the previous chapters this is principally due to two major computational issues: 1) the actual version of the model is unable to reproduce any effect due to TLC 2) SQ may be not correctly implemented in the model. These issues will be examined in details in section 8.4.

8.3 Future studies

A few results reported in previous literature have not been directly analyzed in this dissertation and hence need to be assessed in future researches. In particular it will be necessary to define whether these results require thresholded processing as it has been suggested or rather may be explained in some alternative ways.

First, consider the joint effects of SQ and repetition in reading (Blais & Besner, 2007). The effect of repetition consists in words and nonwords read faster after a single repetition in the task (e.g. Scarborough, Cortese & Scarborough, 1997). This effect is traditionally explained as follow: for words, the lexical entries would retain a higher level of activation when they are repeated than when they are not (see, e.g., Coltheart et al., 2001); instead, for nonwords, the benefit of repetition would arise when the phonological code is translated into an articulatory code (see, e.g., Seidenberg et al., 1996). Blais and Besner (2007) reported that whereas the two factors interact for exception words so that the effect of repetition was larger for degraded than for clear words (a result that, as said, is perfectly consistent with the DRC model's cascaded assumption), the two factors were additive in nonword reading. This finding has been typically interpreted as evidence in favour of an account assuming that activation from the letter level is cascaded to the lexical route, but thresholded to the non-lexical route. We argue, instead, that the pattern of data can be explained by referring to a dual-route account of list composition effects in reading. As said, irregular words were mixed with nonwords in the experiment; so one might expect, for example, that the reliance of the non-lexical route is reduced so to increase lexical contribution in reading (that would be necessary to correctly read irregular words); hence the interaction between SQ and repetition for words. Instead, since the non-lexical route would be de-emphasized in the task, repetition may not influence nonword reading; if this were true we might expect, for example, an interaction between SQ and repetition in nonword reading when nonwords are solely presented in the task. Nevertheless, the crucial point is that there is likely no need to claim the need of a threshold to explain these data since list composition is obviously implicated in these results.

Consider now the joint effects of stimulus quality and regularity in reading aloud, recently examined by Besner et al. (2010). The effect of regularity emerges in languages with inconsistent orthographies like English and consists in slower reading of exception words than regular words (e.g., Seidenberg et al., 1984; Taraban & McClelland, 1987). More specifically, skilled readers are slower to read aloud words like *pint* and *have* because they are exception to the typical relation between spelling and sound in which _INT is pronounced as in *mint* and _AVE is pronounced as in *cave* (see Roberts, Rastle, Coltheart & Besner, 2003). In the DRC framework this effect is explained in terms of competition at the phoneme level; in fact, the lexical and the non-lexical routes would activate different phonemes in the case of an exception word whereas both the routines would activate the same set of phonemes when the word is regular. Hence, when an exception word is presented to be read, the output of the lexical route, which drives the correct pronunciation, would be slowed in the phonemic buffer because of competition from the non-lexical route (which produces a regularization of the input). Besner et al. (2010) reported that when regularity is factorially manipulated with SQ, the two factors interact when only words were presented in the task. However, the effect of degradation was reported to be smaller for irregular than for regular words in their study. Critically, this effect is inconsistent with the DRC model, which simulates the opposite pattern, characterized by the effect of degradation being larger for irregular than for regular words. Moreover, the two factors have been shown to exert additive effects on skilled readers latencies when words where mixed with nonwords in the task, thus providing – according to the authors – evidence against cascaded processing in reading. Besner et al. (2010) explained their results as follow. The interaction obtained when only words were present would depend on the fact that degradation influences the non-lexical route more strongly than the lexical route. The authors showed, in fact, that this interaction is correctly simulated by the DRC model when the strength of the non-lexical route was reduced in degraded condition. Moreover, the additivity obtained when words and nonwords were randomly mixed in the task has been explained by referring again to the lexicalization hypothesis, i.e. processing would be thresholded when both words and nonwords are presented in the task.

We argue that the pattern of results obtained in this study could be interpreted by referring, once again, to list composition effects in degraded reading. First, we note that the effect of degradation is indeed likely to be stronger for the non-lexical route than for the lexical route in the DRC model. This would be due to the fact that, while the lexical route consists of both feedback and feed-forward connections, there is no feedback in the non-lexical route; hence, part of the delay due to degradation would be partially reduced for the lexical route given feedback activation. A few preliminary simulations partially support this hypothesis. In these simulations, regular words were presented to the DRC model and the lexical and the non-lexical routes were alternatively lesioned; it emerged that the effect of degradation for regular words was larger when the lexical route was switched off than when the non-lexical route was switched off; hence, it could be inferred from these data that the effect of SQ is indeed stronger for the non-lexical route than for the lexical route. Moreover, we argue that the balance between the two routes needs to be manipulate to reproduce these results. In fact, an interaction (inverse to that simulated by the DRC) is obtained by Besner et al. (2010) when (regular and irregular) words were presented in the task: since also irregular words were present, one might expect, for example, the lexical contribution being stronger in pure condition; this means either that the lexical route is emphasized or that the non-lexical route is de-emphasized. Consider this second hypothesis: the non-lexical route – which suffers more from degradation – might be weaker in pure list. We argue that a similar account might explain the interpretation proposed by Besner et al. (2010) which, as the authors demonstrated, allows the DRC model to reproduce the data empirically obtained when only words are present in the task. If our hypothesis is plausible, then the additivity of SQ and regularity obtained when also nonwords are presented will be easily explained by assuming, consistently with the proposal expressed in Chapter 6 of this thesis, that the non-lexical route becomes stronger when words and nonwords are randomly mixed in the task compared to when only words are present. So a new job for the feature is to asses these issues through experimental investigation; what is clear, however, is that these data are not a priori against cascaded processing in reading.

8.4 Computational issues

As said, the simulation of the empirical data within the DRC model depends upon two principal computational issues.

The first concerns the Total Letter Confusability: since TLC has an effect for human readers, computational models of reading need to simulate its effects. Hence, we need to reproduce the effect of this variable as well as the pattern of results that depends on TLC in the DRC model.

A second problem regards the simulation of stimulus degradation. Whereas this variable has been largely implemented in computational models of reading we argue that a different solution may be needed at least within the DRC.

8.4.1 Simulating the effect of TLC. A few preliminary results

The Total Letter Confusability is a variable which effect in reading has been recently assessed. The role of this variable in performance of patients affected by pure alexia has been demonstrated, for example, by Fiset et al. (2005). In this dissertation we also showed that unimpaired readers performance is affected by the TLC in degraded presentation: in fact, skilled readers' latencies increase as the TLC of degraded letter strings increases (Chapter 3). Moreover, evidence of sensitivity of human readers to letter confusability for degraded stimuli emerged in the data of the experiment reported in Chapter 4: the correlations between TLC and RTs with letter length partialled out are in fact significant in this experiment for degraded stimuli, $r = .182$, $p = .048$, but not for clear stimuli, $r = -.007$, $p = .94$, consistently with the previous finding.

Since human readers are sensitive to letter confusability we must require computational accounts of reading to be too. In this section we will examine this issue for the DRC model. In fact, if sensitivity to TLC is to be used to simulate in the DRC model the additivities of degradation with other variables in nonword reading, then the DRC model will have to be sensitive to TLC, and this is required more generally because human readers are sensitive to this property of nonwords.

As said, there is no effect of TLC in the DRC model with the current parameters. This is not surprising; the absence of an effect of TLC occurs because the inhibition from feature to letter units is so high relative to their excitation (i.e., .150 vs. .005), a consequence of which is that just one mismatching feature will completely block activation of similar letters. As said, the default values for the Feature-to-Letter Inhibition and Feature-to-Letter Excitation parameters were inherited from the IA model and we don't have to adhere to them. In fact, there is not any reason to assume that the inhibition between feature and letter units is 30 times greater than their excitation. Why shouldn't these parameters be the same?

So we investigated what happens when they are made the same in the newest DRC model's version (the DRC 1.2), by setting both parameters to the value of .005.

When the most confusable letter in DRC's font – the letter *O* – is run with this parameter change, it does activate multiple letters: after 5 processing cycles, there are 17 different letters activated when the input is *O* (the most active letter being *O* itself). When the least DRC-confusable letter *X* is run, there are 4 different letters activated by processing cycle 5 (the most active letter being *X* itself).

What role does this confusability effect have on nonword reading? We used the high-TLC nonword *COLF* (DRC-TLC = 2.7) and the low-TLC nonword *BIDT* (DRC-TLC = 2.1). With the parameter change and degradation (i.e., FLI = FLE = .003), both the nonwords were read aloud in 162 processing cycles; this means that there is no effect of TLC here, even though many more incorrect letters are activated by *COLF* than by *BIDT* (because of the parameter change). It is obvious why there is no TLC effect. Multiple letters activation will only interfere if they can inhibit the correct letter; instead, the default value of Letter-Lateral-Inhibition being used here is zero. Hence we introduced a second parameter exchange by increasing the value of the parameter regulating the lateral inhibition at the letter level from zero to .008. Now the high-TLC nonword *COLF* was read in 182 cycles and the low-TLC nonword *BIDT* in 178 cycles; that is, there's a substantial effect of TLC (4 cycles). Moreover, this effect has been obtained when stimuli were degraded in the simulation; nevertheless, in a non-degraded condition (i.e., FLI = FLE = .005) the low-TLC nonword *BIDT* was read in 146 cycles and the high-TLC nonword *COLF* in 148 cycles; that is, the effect of TLC is much smaller (2 cycles). This means that the effect of TLC is larger when stimuli are degraded than when they are clear, consistently with our empirical data (Chapter 3). As a consequence, the effect of TLC can be correctly simulated by the DRC model when the parameters described above are adequately manipulated.

Now the task is to demonstrate that this way of simulating TLC in the DRC model also shows the pattern of results for length and degradation when TLC is confounded with letter length and when TLC is matched across length. In fact, even it were true that the additivity of degradation and length observed in human reading by Besner and Roberts (2003) occurred because of a confounding between TLC and length, the DRC model still have to be able to simulate this effect, because it is supposed to be sensitive to the same variables that human readers are sensitive to.

To this purpose, we run the Besner and Roberts' (2003) stimuli by applying the parameter modification proposed above in order to make the model sensitive to TLC. As said, short and long nonwords were not matched in terms of TLC in the Besner and Roberts' (2003) experiment: the longer nonwords used in this study had higher TLC values than the shorter nonwords. Moreover, the same is true when the DRC-TLC is considered. Hence, TLC is not matched across letter string length neither for humans nor for the DRC model in this condition.

Mean cycles to criterion are reported in Table 15.

	Stimulus Quality		
	Clear	Degraded	Diff.
Length	Cycles	Cycles	Cycles
Long	157	191.2	34.2
Short	146.9	180.5	33.6
Diff.	10.1	10.7	

Table 15. Mean cycles for the Besner and Roberts' (2003) stimuli (TLC confounded across length).

Here the effect of letter string length is not much different for clear and degraded stimuli, consistently with the additivity reported by Besner and Roberts (2003). Analysis showed that – if anything – the length effect is smaller for clear than for degraded stimuli[38], $F(1,62) = 3.8$, $MSE = .829$, $p = .057$, consistently to what we found in our experiment (Chapter 4).

[38] Even it were true that the longer nonwords have higher TLC values than the shorter nonwords in the Besner and Roberts' (2003) study both for human readers and for the DRC model, the specific values of TLC for humans clearly differ from the specific DRC-TLC values for these stimuli. As a consequence, it is not surprising that the results obtained in this simulation only partially reproduce the empirical data documented by the authors.

Now we need to show that when short and long nonwords are matched on TLC, there is an interaction between SQ and letter string length with the length effect smaller for degraded stimuli. In order to test this prediction we need to create a specific set of stimuli, so to have short and long nonwords matched on the DRC-TLC values. We would have liked to choose nonwords that were 3, 4, 5 and 6 letters long with TLC matched across the four conditions, but this is not possible with DRC model's letter confusabilities. The difference between the highest and lowest confusability values is not large enough to have equal TLC for 3 letter and 6 letter strings. So we used 3 and 4 letter nonwords matched on TLC, and separately 5 and 6 letter nonwords matched on TLC; 32 items for each of the 4 cells matched for their initial phoneme have been selected.

Mean cycles to criterion are reported in Table 16 and in Table 17.

	Stimulus Quality		
	Clear	Degraded	Diff.
Length	Cycles	Cycles	Cycles
4 letters	144.4	175.9	31.5
3 letters	142.2	175.5	33.3
	2.2	0.4	

Table 16. Mean cycles for 3 and 4 letter nonwords matched on TLC in function of SQ.

	Stimulus Quality		
	Clear	Degraded	Diff.
Length	Cycles	Cycles	Cycles
6 letters	156.1	189.6	33.5
5 letters	153.8	189.2	35.4
	2.3	0.4	

Table 17. Mean cycles for 5 and 6 letter nonwords matched on TLC in function of SQ.

When short and long nonwords are matched in terms of TLC, the effect of degradation is smaller for degraded than for clear stimuli. Analysis showed a highly significant interaction between SQ and letter string length both for 3 and 4 letter nonwords, $F(1,62) = 23.1$, $MSE = 1.3$, $p < .001$, and for 5 and 6 letter nonwords, $F(1,62) = 9.8$, $MSE = .829$, $p < .005$.

To summarize, we showed that the DRC model perfectly reproduces the patter of empirical data by adjusting the values of the parameters that regulate the model's sensitivity to letter similarity. In particular, when the parameter that regulates the inhibition between the features and the letters is reduced so to match the value of the parameter regulating their excitation and letter lateral inhibition is implemented in the model (consistently with the theoretical assumption), the following results are obtained:

1. both human readers and the DRC model show an effect of TLC, which is larger when stimuli are degraded than when stimuli are clear;

2. both human readers and the DRC model show an interaction between degradation and length on nonword reading, with the length effect smaller for degraded than for clear stimuli when TLC is matched across the different values of nonword length;

3. both human readers and the DRC model show additivity of the effects of degradation and length on nonword reading RTs (or a smaller effect of length for clear than for degraded stimuli) when TLC is confounded with length.

Our preliminary work thus suggests that the DRC model is successful in predicting and simulating the results depending on TLC. Now a job for the future is to see whether the DRC 1.2 with the new parameter setting can simulate all the effects that its actual version and the previous versions could. Moreover, now that we have a way of simulating TLC effects, it could be possible to simulate letter-by-letter reading with the DRC model and in particular to simulate the various results depending on this variable.

Obviously, a further challenge for feature researches will be also determining whether and how computational models of reading (besides the DRC model) can simulate the effects due to letter confusability as well as the patter of results depending on this variable.

8.4.2 Is degradation correctly implemented? A different proposal

It is largely assumed that degradation can be implemented in the DRC model by reducing the connection weights between the feature and the letter units. However, stimulus quality should instead affect the rate of activation gain at the visual feature level, since "*Visually degrading the stimulus will have the effect of lowering the effectiveness of the stimulus in activating all of the relevant feature detectors*" (McClelland, 1979, p. 292).

Hence, we argue that SQ is actually not correctly implemented in the DRC model.

As said, in the current computational version of the model, a visual feature can be only on or off; as a consequence, the activation accrual at the feature level cannot be affected by degradation. If it were possible to slow the rate of processing at the visual feature level, it would be only by adding a constant of time to performance. Simulation in fact begins when all the feature units reached an activation of 1.0; feature units are clamped and reach an activation of 1.0 in one cycle. In real life, however, the activation in these units will not go from 0 to 1.0 instantaneously but this process will take a certain time. A way to implement degradation in the DRC model might thus be by delaying the time that feature units need to reach threshold. However, the effect of this manipulation will be simply to add a time constant to performance. Hence, the model will be incapable of producing anything else than additive effects of SQ and a second factor affecting one of the subsequent levels assumed in the model. Since interactive effects of these types of variables have been instead largely documented (e.g., Blais & Besner, 2007; Ferguson et al., 2009; O'Malley & Besner, 2008; Yap & Balota, 2007; our results reported in Chapter 4 and 5) this cannot be an adequate solution.

Given that activation is cascaded in the model (i.e., a change in rate of activation in early processing units will change the rate of activation downstream), it has been proposed that a reduction of activation at the feature level may be modelled by reducing the weights of the connections between the feature and the letter units. In other words, the effect of a reduction in the asymptotic level of activation of the feature units may be reproduced by reducing the rate at which activation accrues at the letter level. Hence, degradation is actually implemented in the DRC model by reducing the parameters regulating excitation and inhibition form the features to the letters. It is clear, however, that this solution has been required by the current computational architecture of the model; nevertheless, this implementation does not reflect any theoretical issue nor the effect that SQ has on human performance.

We argue that a more adequate implementation may be possible by changing the computational version of the DRC model as follow. As said, *"individual features (...) are set to on or off (1 or 0)"* (Coltheart et al., 2001, p. 213) in the current program and this organization was inherit from the progenitor of the DRC, the IA model; as McClelland and Rumelhart (1981) stated, in fact, *"It is assumed that features are binary and that we can extract either the presence or the absence of a particular visual feature"* (p. 381). In short, the feature level is actually not cascaded, but rather thresholded. This organization, however, is used purely for convenience and doesn't rely on any specific theoretical assumption. On the contrary, cascaded processing is assumed to occur in the system. Hence, the idea might be to allow activation to cascade from the visual feature level to the subsequent letter level. This might be obtained by allowing a visual feature to accumulate activation continuously up to its asymptotic value, the same solution that is actually implemented in the other levels of the model; in other words, activation in each feature unit would grow over time in the continuum 0-1.0 and the level of activation in a precise moment in time would depend on the quality of the stimulus in input, following the assumption that the rate at which activation rises at the feature level depends on SQ, i.e. activation rises faster for clear stimuli, slower for degraded stimuli. As a consequence, whereas features corresponding to clear letters will be fully activated when the stimulus is presented, features corresponding to degraded letters will be only partially activated by the stimulus in input; in turn, degraded letters will be less activated than clear letters and responses will be thus slower for degraded letter strings.

We argue that modifying the model in this way may allow the simulation of the effects that SQ and TLC have with lexical variables. We argue, in fact, that the pattern of results obtained in this context may depend on the different levels of processing at which these variables have their effects; specifically, we suggested that the additivities of SQ and word frequency/lexicality obtained when words and nonwords were mixed in the task could depend on the indirect effects that the activation from the orthographic lexicon has on SQ that, in turn, would depend on the fact that there is no feedback from the letter to the feature units. Clearly, implementing degradation within the feature level will be necessary to correctly simulate these effects. Moreover, we argue that such an implementation has to be realized in the DRC model not only to allow it to correctly simulate the effect of SQ, but also to fully implement the model's theoretical commitment to cascaded processing.

8.5 Thinking to a threshold

Before coming to a general conclusion, a few points need to be further discussed.

As largely asserted in this dissertation, our data clearly deny the hypothesis of thresholded processing at the letter level of the DRC model. But is there a threshold in any other level assumed in the model? In general we argue it is not. However, consider how the response is usually generated by participants in a reading task. Typically, in reaction time experiments, subjects are instructed to respond as rapidly as possible maintaining a high level of accuracy. How do subjects decide that the time has come to initiate a response (see McClelland, 1979, p. 304)? In traditional discrete stage models, it is relatively easy to understand when participants would release they responses since the output of processing is all-or-nothing and becomes thus available at some particular instant in time: after that time, the correct response can be executed; before it, subjects would simply have to guess. However, in terms of cascaded accounts, there is no a specific instant in time before which responding would be at chance and after which it would be correct, since activation continuously increases in processing units gradually leveling off at some maximal level. In such a situation, what does the instruction to respond as rapidly and accurately as possible mean? One possibility is that participants set an implicit deadline consistent with a low enough error rate (Ollman, 1977); this strategy, however, would not explain why RTs differ for different experimental conditions, even when the items representing these conditions are mixed within the same block of trials. An alternative is that subjects adopt an activation criterion and respond when activation in a response unit reaches a level that is sufficient to ensure an acceptably low error rate (Grice, 1968). A similar solution is implemented in the DRC model: in fact "*the model is considered to have determined the pronunciation of a monosyllabic letter string when it has been activated (to some criterion of satisfaction) all of the phonemes of that letter string*" (Coltheart et al., 2001, p. 217); in other words, pronunciation occurs in the DRC model when all the phonemes reach a threshold. However, activation flows in a cascaded fashion through the reading system and this mechanism only allows the execution of responses.

A related issue regards the DRC model's non-lexical route and, in particular, how the (serial) movement from left to right is implemented in the model. In the last computational version of the model (i.e., the DRC 1.2), the non-lexical route moves on the next letter when the currently being activated phoneme reaches a critical level of activation: evidently this could be described as thresholding the phoneme level. Again, this does not mean that processing is strictly thresholded, but rather that a thresholded mechanism is implemented in the model in order to represent the spatially serial processing assumed within the non-lexical route.

To conclude, we argue that processing in the reading system is cascaded but a thresholded solution is implemented at the phoneme level of the DRC model 1) to allow pronunciation 2) to implement the serial movement of the non-lexical route.

The second issue we need to point out can be expressed as follow: is thresholded processing needed to explain additive effects of variables? Within discrete stage models, additive effects of variables are easily explained by assuming that those variables influence different levels of processing; an interaction between variables would instead mean that those variables affect the same level of processing. Conversely, this logic does not apply to cascaded models: in fact, if activation cascades through the levels, also variables affecting different levels of processing can interact one another. It follows that, whereas factorial manipulations may provide useful information to delineate processing sequence in reading according to discrete stage models, this interpretation is not longer valid within cascaded accounts.

Besides this crucial issue, a question remains to be dealt with: can additive effects of variables be explained within a cascaded framework? This topic is extremely relevant in researches on visual word recognition: in fact, if only interactions are predicted by cascaded processing, then this account has probably to be rejected; it would be indeed unrealistic to assume that factors manipulated in the task can only interact with one another. To better understand this point assume an experiment in which two variables affecting different levels of processing are jointly manipulated: call these variables A and B and assume that A influences an earlier process than B. In a cascaded model, the effect of the variable A is not resolved within the level which A affects; on the contrary, the effect of the variable A will influence processing downstream in the system. We argue, however, that this does not necessarily mean that A and B interact: the variable A could simply have an effect on performance (affecting processing downstream in the system) without interacting with the effect of the variable B. In other words, we suggest that the effect of an early factor may cascade to the subsequent levels of processing still being additive with the effects of variables operating at those levels. The situation, however, is likely to change when cascaded models also assume interactive activation. In these circumstances, in fact, not only the effect of A cascades to the level influenced by the variable B, but the effect of B also feeds back to the level affected by A; since feedback from a later level contributes activation at the earlier levels, the two variables are likely to interact.

To summarize, we are inclined to believe that cascaded processing is consistent with additive effects of variables, whereas cascaded models which also assume interactive activation may not be able to reproduce these effects. Whether this interpretation is valid remains to be fully determined. It seems to us, however, that the results discussed in the present dissertation provide partial support to this interpretation. Some of the interactions we reported are in fact explained in terms of feedback activation rather than by cascaded processing per se[39]: in many circumstances the two factors interact because the variable having a later effect has a role on the earlier variable. To use an example, consider the interaction between SQ and orthographic neighbourhood size (N) in nonword reading (see Chapter 5): it is not just because the delay in processing caused by degradation cascades to the lexical level that the effect of N increases in degraded reading; rather, the feedback from the lexicon contributes activation at the letter level thus reducing the effect of degradation for nonwords with orthographic neighbours and hence producing a larger effect of N in degraded condition.

[39] Interactive activation makes of course sense only within cascaded frameworks. In discrete stage models, in fact, there would be not any effect if activation feeds back from later levels to a level which processing is already ended.

If our reasoning is plausible, it follows that even processing may in fact be cascaded in visual word recognition, perhaps not all the levels of processing assumed in the reading system may communicate with one another through feedback connections. At this purpose consider, for example, the DRC model: as said, there is no feedback in the model from the letter to the feature units. We suggested in this dissertation that this organization may indeed be crucial to explain the additivities of SQ and lexical variables observed when words and nonwords are mixed in the task (see Chapter 6). Regardless of our specific proposal, this feature could in fact allow the DRC model to account for additive effects of a variable influencing the visual feature analysis (e.g., SQ) and some other factors influencing later processes. Unfortunately, it has been not yet determined whether the whole patter of data could be simulated by a similar architecture. It seems to us to be critical, however, that additive effects of variables documented in previous literature always involve degradation as one of those variables; in other words, to the best of our knowledge, additive effects have never been reported when the factors manipulated in the task affect levels of processing which are subsequent to the visual feature analysis. Evidently, this may be a strong argument in favour of our account. Clearly, defining how visual word recognition needs to be modelled in order to reproduce additive and interactive effects of factors will be a critical issue for future researches.

Finally, it has to be noted that cascaded and discrete stage models are not the solely available accounts. On the contrary, an intermediate position also exists, even if computational models of reading and visual word recognition have typically ignored this idea.

Consider first the following question: what do we mean, exactly, with cascaded processing? Assuming cascaded processing does not mean to assume that subjects are able to identify, for example, the letters in a stimulus without the result of the feature analysis; as a consequence, the logical requirement of the task itself requires that at least some of the processes involved in performance occur in a strict succession. Cascaded processing more likely means to reject the idea of traditional discrete stage models that one component of processing must be completed before a second can start; hence, according to cascaded models, even when a process depends on the output of another process, the later one will start *before* the previous is completely ended. In order to have a better understanding of this point, consider the formulation suggested by Norman and Bobrow (1975). The authors hypothesized that the output of each process could be a set of quantities, each one indicating the degree of confidence that one of the several possible conclusions about the stimulus in input is correct. For example, at some instant in time, the output of the feature analysis might indicate a 20% chance that there is a vertical line on the left of the pattern in input and a 5% chance that there is a horizontal line across the middle; a bit later, the same outputs might indicate values of 35% and 60%. According to cascaded processing, the outputs from the feature level are always available and the process of letter identification would be using this changing information over time.

Now, imagine to apply this same logic to thresholded processing. In traditional stage models, a threshold is reached in a processing unit when activation in that unit reaches its maximum value, which means that information processing occurring in that unit is completely ended; to exemplify, according to traditional discrete stage models, activation would be passed to the letter level when the output of the feature analysis indicates a 100% chance that the stimulus in input has those visual features. As said, several empirical data contrasting this hypothesis exist. A different interpretation is however possible by assuming that the threshold does not reflect the maximum value of activation in the processing units; rather, a similar thresholded mechanism may be set to an amount of activation that is considered satisfactory in that unit: once this critical level is reached, then activation would flow to the subsequent level in a purely cascaded fashion. In other words, it is possible that processing occurring at an early level needs to collect a certain amount of information about the stimulus in input before cascading to a subsequent level. Using the previous example, it could be that activation is passed on to the letter level only when the output of the feature level analysis indicates with a certain chance (say, e.g., 30%) the presence of some visual features in input. A possible reason for a similar organization may be a principle of cognitive ergonomic; one may argue, in fact, that there would be no reasons to pass information on to latter stages when the degree of confidence about some characteristics of the stimulus in input is very low: such a strategy would be in fact highly demanding in terms of cognitive resources and perhaps counter-productive for performance. Whether this proposal fits the empirical evidence needs to be evaluated. It seems to us, however, that such an account could be a straightforward solution to reproduce additive and interactive effects of variables, perhaps being a valid alternative to a purely cascaded account. It may be hence interesting to implement a similar information processing modality in future computational modelling of visual word recognition and reading aloud.

8.6 General conclusion

The assumption of cascaded processing is central in the DRC model and in many other models of language processing following the work by McClelland and Rumelhart (1981). Previous simulation works, considered in the light of results from skilled readers, demonstrated that the assumption of cascaded processing is problematic. These studies concluded that, since the DRC model fails to account for the effects showed by human readers, it requires a modification; in general, a threshold at the letter level has been introduced as a computational solution that would allow the model to simulate the empirical data obtained by varying stimulus quality in conjunction with lexical and non-lexical factors in reading aloud tasks.

The main goal of the empirical activities reported in this thesis has been to test the hypothesis of thresholded processing in the reading system. First of all, we collected empirical evidence demonstrating that processing in reading has to be cascaded. Moreover, we gave an explanation of the previous apparently problematic results. From one hand, we showed that the additivities of SQ and another variable observed in nonword reading arose because of a confounding with the Total Letter Confusability, a variable that is involved in reading of degraded letter strings; also, we demonstrated that the DRC model can in fact simulate these empirical data simply adjusting the values of a few parameters that regulate its sensibility to letter similarity. From the other hand, we argue that the additivities of SQ and lexical variables obtained when words and nonwords are randomly mixed in the task can be explained by considering list composition effects in degraded reading; we also suggested that degradation should be differently implemented in the DRC model in order to attempt the simulation of these results. In short, we demonstrated that the additivities previously reported do not necessarily claim the need of a threshold in the reading system; on the contrary, these data can be explained by modelling reading with a dual-route account assuming cascaded processing.

As a consequence, the answer to the critical question investigated in this thesis – *does the Dual-Route Cascaded model require a threshold*? – is clearly *No*. Rather, the computational version of the model needs to be only partially modified in order to reproduce the effects due to letter confusability and to correctly implement stimulus degradation. More generally, the data obtained so far indicate that there is likely to be no thresholded processing in the reading system; factorial manipulations of factors will be extremely useful to further analyze this issue.

Finally, practical and theoretical implications of the empirical work presented in this thesis are evident. From one hand, our data showed that the DRC model of reading does not require any radical modification involving its cascaded nature. From the other hand, the researches that have been presented sustain a cascaded framework as the account that – (at least) at present – better models the cognitive processes underling reading aloud and visual word recognition.

References

Adams, M. J. (1979). Models of word recognition. *Cognitive Psychology*, **11**, 133-176.

Andrews, S. (1997). The effect of orthographic similarity on lexical retrieval: Resolving neighbourhood conflicts. *Psychonomic Bulletin & Review*, **4**, 439-461.

Andrews, S. (1989). Frequency and neighborhood effects on lexical access : Activation or search ? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **15**, 802-814.

Andrews, S., & Scarratt, D. R. (1998). Rule and analogy mechanisms in reading aloud nonwords: Hough dou peapel rede gnew wirds? *Journal of experimental psychology: human perception and performance*, **24**, 1052-086.

Arduino, L. S., & Burani, C. (2004). Neighborhood Effects on Nonword Visual Processing in a Language with Shallow Orthography. *Journal of Psycholinguistic Research*, **33** (1), 75-95.

Arguin, M., & Bub, D. (2005). Parallel processing blocked by letter similarity in letter dyslexia: a replication. *Cognitive Neuropsychology*, **22**, 589-602.

Arguin, M., Fiset, S., & Bub, D. (2002). Sequential and parallel letter processing in letter-by-letter reading. *Cognitive neuropsychology*, **19**, 535-555.

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R.* Cambridge University Press.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modelling with crossed random effects for subjects and items. *Journal of Memory and Language*, **59**, 390-412.

Baayen, R., H., Piepenbrock, R., & van Rijn, H. (1993). *The CELEX lexical database* (CD/ROM). Linguistic Data Consortium, University of Pensilvania.

Balota, D. A., & Abrams, D. (1995). Mental chronometry: Beyond onset latencies in the lexical decision task. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **21**, 1289-1302.

Balota, D., Cortese, M. J., Sergen-Marshall, S. D., Spieler, D. H., & Yap, M. J. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General*, **133**, 283-316.

Basso, A., Taborelli, A., & Vignolo, L. A. (1978). Dissociated disorders of speaking and writing in aphasia. *Journal of Neurology, Neurosurgery and Psychiatry*, **42**, 1115-1124.

Bates, D., Maechler, M., & Dai, B. (2008). lme4: *Linear mixed-effects models using S4 classes. R package version 0.999375-28*. http://lme4.r-forge.r-project.org/

Becker, C. A., (1979). Semantic context and word frequency effects in visual word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, **5**, 252-259.

Besner, D., & O'Malley, S. (2009). Additivity of factor effects in reading tasks is still a challenge for computational models: reply to Ziegler, Perry, and Zorzi (2009). *Journal of Experimental Psyhcology: Learning, Memory, and Cognition*, **35** (1), 312-316.

Besner, D., O'Malley, S., & Robidoux, S. (2010). On the joint effect of stimulus quality, regularity and lexicality when reading aloud: New challenges. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **36** (3), 750-764.

Besner, D., Reynolds, M., & Chang, K. (2003). Basic processes in reading: Evolution of the Dual Route Cascaded model. *Paper presented at the 13th meeting of the Canadian Society for Brain, Behaviour and Cognitive Science*, Hamilton, ON.

Besner, D., Reynolds, M., & O'Malley, S. (2009). When underadditivity of factor effects in the Psychological Refractory Period paradigm implies a bottleneck: Evidence from psycholinguistic. *The Quarterly Journal of Experimental Psychology*, **62** (11), 2222-2234.

Besner, D., & Roberts, M.A. (2003). Reading nonwords aloud: Results requiring change in the dual route cascaded model. *Psychonomic Bulletin & Review*, **10** (2), 398-404.

Besner, D., & Smith, M. (1992). Models of visual word recognition: When obscuring the stimulus yelded a clearer view. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **18**, 468-482.

Besner, D., & Swan, M. (1982). Models of lexical access in visual word recognition. *Quarterly Journal of Experimental Psychology*, **34A**, 313-325.

Beversdorf, D. Q., Ratcliffe, N. R., Rhodes, C. H., & Reeves, A. G. (1997). Pure alexia – clinical-pathologic evidence for a lateralized visual language association cortex. *Clinical Neurophathology*, **16**, 328-331.

Binder, J. R., & Mohr, J. P. (1992). The topography of callosal reading pathways. A case-control analysis. *Brain*, **115**, 1807-1826.

Blais, C., & Besner, D. (2007). Reading Aloud. When the Effect of stimulus Quality Distinguishes Between Cascaded and Thresholded Components. *Experimental Psychology*, **54** (3), 215-224.

Borowsky, R., & Besner, D. (1993). Visual word recognition: A multistage activation model. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **19**, 813-840.

Bouma, H. (1973). Visual interference in the parafoveal recognition of initial and final letters of words. *Vision Research*, **13**, 767-782.

Bramwell, B. (1897). Illustrative cases of aphasia. *The Lancet*, **i**, 1256-1259 (Reprinted in Cognitive Neuropsychology, 1984, I, 249-258).

Brown, G. D. A. (1987). Resolving inconsistency: A computational model of word naming. *Journal of Memory and Language*, **26**, 1-23.

Clarke, R, & Morton, J. (1983). Cross modality facilitation in tachistoscopic word recognition. *Quarterly Journal of Experimental Psychology*, **35A**, 79-96.

Chateau, D., & Lupker, S. J. (2003). Strategic effects in word naming: examining the Route-Emphasis versus Time-Criterion accounts. *Journal of Experimental Psychology: Human Perception and Performance*, **29** (1), 139-151.

Coltheart, M. (1978). Lexical access in simple reading tasks. In G. Underwood (Ed.), *Strategies of information processing* (pp. 112-174). New York: Academic Press.

Coltheart, M. (Ed.) (1998). *Pure alexia (letter-by-letter reading)*. Hove, England: Psychology Press.

Coltheart, M. (1999). Modularity and cognition. *Trends in Cognitive Sciences*, **3**, 115-120.

Coltheart, M. (2004). Are there lexicons? *The Quarterly Journal of Experimental Psychology*, **57A** (7), 1153-1171.

Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of reading aloud : Dual-route and parallel-distributed-processing approaches. *Psychological Review*, **100**, 589-608.

Coltheart, M., Davelaar, E., Jonasson, J., e Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and performance IV* (pp. 535-555). New York: Accademic Press.

Coltheart, M. & Rastle, K., (1994). Serial processing in reading aloud: Evidence for dual-route models of reading. *Journal of Experimental Psychology: Human Perception and Performance*, **20**, 1197-1211.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A Dual Route Cascaded model of visual word recognition and reading aloud. *Psychological Review*, ***108***, 204 – 256.

Courrieu, P., Farioli, F., & Grainger, J. (2004). Inverse discrimination time as a perceptual distance for alphabetic characters. *Visual Cognition*, **11**, 901-919.

Damasio, A. R., & Damasio, H. (1983). The anatomic basis of pure alexia. *Neurology*, **33**, 1573-1583.

Déjerine, J. (1982). Contribution à l'étude anatomo-pathologique et clinique des différentes variétés de cécité verbale. *Compte Rendu Hebdomadaire des Séances et Mémoires de la Société de Biologie*, **4**, 61-90.

Donders, F. C., (1868-1869). Over de snelheid van psychische processen. Onderzoekingen gedann in het Physiologish Laboratorium der Utrechtsche Hoogeshool, Tweede Reeks, II, 92-120. In W. G. Koster (Ed. and transl.), *Attention and performance II*. Amsterdam: North-Holland, 1969. (Reprinted from *Acta Psychologica*, 1969, **30**, 412-431).

Ellis, A. W., Young, A., & Anderson, C. (1988). Modes of word recognition in the left and right cerebral hemisphere. *Brain and Language*, **35**, 254-273.

Estes, W. K. (1975). The locus of inferential and perceptual processes in letter identification. *Journal of Experimental Psychology: General*, **1**, 122-145.

Estes, W. K., Allemeyer, D. H., & Reder, S. M. (1976). Serial position functions for letter identification at brief and extended exposure durations. *Perception & Psychophysics*, **19**, 1-15.

Ferguson, R., Robidoux, S., & Besner, D. (2009). Reading aloud: Evidence for contextual control over lexical activation. *Journal of Experimental Psychology: Human Perception and Performance*, **35**, 499-507.

Fiset, D., Arguin, M., Bub, D., Humphreys, G.W., & Riddoch, M.J. (2005). How to make the word-length effect disappear in letter-by-letter dyslexia. *Psychological Science*, **16** (7), 535-541.

Fiset, S., Arguin, M., & Fiset, D. (2006). An attempt to simulate letter-by-letter dyslexia in normal readers. *Brain and Language*, **98**, 251-263.

Fiset, D., Arguin, M., & McCabe, E. (2006). The breakdown of parallel letter processing in letter-by-letter dyslexia. *Cognitive Neuropsychology*, **23**, 240-260.

Fiset, D., Gosselin, F., Blais, C., & Arguin, M. (2006). Inducing letter-by-letter dyslexia in normal readers. *Journal of Cognitive Neuroscience*, **18** (9), 1466-1476.

Forster, K. I., & Taft, M. (1994). Bodies, antibodies, and neighborhood-density effect in masked form priming. *Journal of Experimental Psychology: Learning, memory, and Cognition*, **20**, 844-863.

Gilmore, G.C., Hersh, H., Caramazza, A., & Griffin, J. (1979). Multidimensional letter similarity derived from recognition errors. *Perception and Psychophysics*, **25**, 425-431.

Glushko, R. J. (1979). The organization and activation of orthographic knowledge in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance,* **5**, 674-691.

Grainger, J., & Jacobs, A. M. (1996). Orthographic processing in visual word recognition: A multiple read-out model. *Psychological Review*, **103**, 518-565.

Grainger, J., Rey, A., & Dufau, S. (2008). Letter perception: from pixels to pandemonium. *Trends in Cognitive Sciences*, **12** (10), 381-387.

Grice, G. R. (1968). Stimulus intensity and response evocation. *Psychological Review*, **75**, 359-373.

Harm, M. W., & Seidenberg, M. S. (1999). Phonology, reading acquisition, and dyslexia: Insights from connectionist models. *Psychological Review,***106**, 491–528.

Herdman, C. M. (1992). Attentional resource demands of visual word recognition in naming and lexical decisions. *Journal of Experimental Psychology: Human Perception and Performance*, **15**, 124-132.

Howard, D., & Franklin, S. (1988). *Missing the meaning? A cognitive neuropsychological study of the processing of words by an aphasic patient*. Cambridge, MA: MIT Press.

Huey, E. B. (1908). *The psychology and pedagogy of reading*. Macmillan: New York.

Jacobs, A. M., & Grainger, J. (1994). Models of visual word recognition: Sampling the state of the art. *Journal of Experimental Psychology: Human Perception and Performance*, **20**, 1311-1334.

Jared, D. (1997). Spelling-sound consistency affects the naming of high-frequent words. *Journal of Memory and Language*, **36**, 505-529.

Jared, D. (2002). Spelling-sound consistency and regularity effects in word naming. *Journal of Memory and Language*, **46**, 723-750.
Jared, D., McRae, K., & Seidenberg, M. S. (1990). The basis of consistency effects in word naming. *Journal of Memory and Language*, **29**, 687-715.
Job, R., Peressotti, F., & Cusinato, A. (1998). Lexical effects in naming pseudowords in shallow orthographies: Further empirical data. *Journal of Experimental Psychology: Human Perception and Performance*, **24**, 622-630.
Johnston, J. C., McCann, R. S., & Remington, R. W. (1995). Chronometric evidence for two types of attention. *Psychological Science*, **6**, 365-369.
Johnston J. C., & McClelland, J. L. (1980). Experimental texts of hierarchical model of word identification. *Journal of Verbal Learning and Verbal Behaviour*, **19**, 503-524.
Kang, S. H. K., Balota D. A., & Yap, M. J. (2009). Pathway control in visual word processing: Converging evidence from recognition memory. *Psychonomic Bulletin & Review*, **16** (4), 692-698.
Kinoshita, S., & Lupker, S. J. (2003). Priming and attentional control of lexical and sublexical pathways in naming: A reevaluation. *Journal of Experimental Psychology: Learning, Memory and Cognition*, **29**, 405-415.
Kinoshita, S., & Lupker, S. J. (2007). Switch costs when reading aloud words and nonwords: Evidence for shifting route emphasis? *Psychonimic Bulletin & Review*, **14**, 449-454.
Kinoshita, S., Lupker, S. J., & Rastle, K. (2004). Modulation of regularity and lexical effects in reading aloud. *Memory and Cognition*, **32** (8), 1255-1264.
Kwantes, P. J., & Marmurek, H. C. (2007). Controlling lexical contributions to the reading of pseudohomophones. *Psychonomic Bulletin and Review*, **14**, 377-378.
LaBerge, D., & Samuels, S. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology*, **6**, 293-323.
Lhermitte, F., & Derouesné, J. (1974). Paraphasies et jargonaphasies dans le langage oral avec conservation du langage écrit. *Revue Neuropsychology*, **130**, 21-38.
Loomis, J.M. (1982). Analysis of tactile and visual confusion matrices. *Perception and Psychophysics*, **31**, 41-52.
Lupker, S. J., Brown, P., & Colombo, L. (1997). Strategic control in naming task: changing routes or changing deadlines? *Journal of Experimental Psychology: Learning, Memory and Cognition*, **23** (3), 570-590.
McCann, R. S., & Besner, D. (1987). Reading pseudohomophones: Implications for models of pronunciation assembly and the locus of word-frequency effects in naming. *Journal of Experimental Psychology: Human Perception and Performance*, **13**, 14-24.
McClelland, J. L. (1976). Preliminary letter identification in the perception of words and nonwords. *Journal of Experimental Psychology: Human Perception and Performance,* **1**, 80-91.
McClelland, J.L. (1979). On the time relations of mental processes: An examination of system of processes in cascade. *Psychological Review*, **86**, 287-330.
McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: 1. An account of the basic findings. *Psychological Review,* **88**, 357-407.
Melville, J. P. (1957). Word-length as a factor in differential recognition. *American Journal of Psychology*, **70**, 316-318.
Meyer, D. E., Schvaneveldt, R. W., & Ruddy, M. G. (1975). Loci of contextual effects on visual word recognition. In P. M. A. Rabbitt & S. Dornic (Eds.) *Attention and Performance V* (pp. 98-118). San Diego: Academic Press.
Monsell, S., Patterson, K. E., Graham, A., Hughes, C. H., & Milroy, R. (1992). Lexical and sublexical translation of spelling to sound: Strategic anticipation of lexical status. *Journal of Experimental Psychology: Learning, Memory and Cognition,* **18**, 452-467.
Morton, J. (1969). Interaction of information in word recognition. *Psychological Review,* **76**, 165-178.

Morton, J. (1979). Facilitation in word recognition: Experiments causing change in the logogen model. In P. A. Kolers, M. E. Wrolstad, & H. Bouma (Eds.), *Processing of visual language: 1*. (pp. 259-268). New York: Plenum Press.
Morton, J. (1980). The logogen model and orthographic structure. In U. Frith (Ed.), *Cognitive processes in spelling.* London: Academic Press.
Morton, J., & Patterson, K.E. (1980). A new attempt at an interpretation, or, an attempt at a new interpretation. In M. Coltheart, K. Patterson, & J.C. Marshall (Eds.), *Deep dyslexia* (pp. 91-118). London: Routledge and Kegan Paul.
Mousikou, B., Coltheart, M., Saunders, S., & Yen, L. (2010). Is the orthographic/phonological onset a single unit in reading aloud? *Journal of Experimental Psychology: Human Perception & Performance*, **36**, 175-194.
Mozer, M. C. (1987). Early parallel processing in reading: A connectionist approach. In M. Coltheart (Ed.), *Attention and performance XII: The psychology of reading* (pp. 83-104). London: Erlbaum.
Mulatti, C. (2005). *Modelling computational modelling of reading*. Unpublished doctoral dissertation, Università degli Studi di Padova.
Mulatti, C., & Job, R. (2003a). *Lettura di parole isolate: un simulatore, un esperimento e una simulazione*. Laboratorio di Scienze Cognitive, Università degli studi di Trento, Report No. 19, Rovereto.
Mulatti, C., & Job, R. (2003b). Lettura di parole "straniere" in italiano: effetto della regolarità e della posizione dell'irregolarità. *Giornale Italiano di Psicologia*, **30**, 883-893.
Mulatti, C. & Job, R. (2004). Implementing the DRC model of reading aloud in Italian. *International Journal of Psychology*, **39**, 5-6.
Mulatti C., Peressotti F., & Job R. (2007). Reazing and zeading: Which is faster? The position of the diverging letter in a pseudoword determines reading time. *The Quarterly Journal of Experimental Psychology*, **60**, 1005-1014.
Mulatti, C., Reynolds M. G., & Besner, D. (2006). Neighborhood effects in reading aloud: new findings and new challenges for computational models. *Journal of Experimental Psychology: Human Perception and Performance*, **32** (4), 799-810.
Navon, D., & Miller, J. (2002). Queuing or sharing? A critical evaluation of the single bottleneck notion. *Cognitive Psychology*, **44**, 193-251.
Neisser, U. (1967). *Cognitive Psychology*. Appleton-Century-Crofts: New York.
Norman, D. A., & Bobrow, D. G. (1975). On the data-limited and resources-limited processes. *Cognitive Psychology*, **7**, 44-64.
O’Malley, S., & Besner, D. (2008). Reading Aloud: Qualitative Differences in the Relation Between Stimulus Quality and Word Frequency as a Function of Context. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **34** (6), 1400-1411.
O’Malley, S., Reynolds, M.G., & Besner, D. (2007). Qualitative differences between the joint effects of stimulus quality and word frequency in reading aloud and lexical decision: Extension to Yap e Balota. *Journal of Experimental Psychology: Learning Memory and Cognition*, **33**, 451-458.
Ollman, R. (1977). Choice reaction time and the problem of distinguishing task effects from strategy effects. In S. Dornic (Ed.), *Attention and performance VI*. Hillsdale, N.J.: Erlbaum.
Paap, K. R., & Noel, R. W. (1991). Dual-route models of print to sound: Still a good horse race. *Psychological Research/Psychologische Forschung*, **53**, 13-24.
Pashler, H. E. (1984). Processing stages in overlapping tasks: Evidence for a central bottleneck. Journal *of Experimental Psychology: Human Perception and Performance*, **10**, 358-377.
Pashler, H. E. (1994). Dual task interference in simple tasks: Data and theory. *Psychological Bulletin*, **116**, 220-244.
Patterson, K. E., & Kay, J. (1982). Letter-by-letter reading: psychological description of a neurological syndrome. *Quarterly Journal of Experimental Psychology*, **34A**, 411-441.

Patterson, K. E., & Morton, J. (1985). From orthography to phonology: An attempt at an old interpretation. In K. E. Patterson, J . C . Marshall, & M. Coltheart (Eds.), *Surface dyslexica: Neuropsychological and cognitive studies of phonological reading* (pp. 335-359). London: Erlbaum.

Peereman, R., & Content, A. (1995). Neighborhood size effect in naming: Lexical activation or sub-lexical correspondence? *Journal of Experimental Psychology: Learning Memory and Cognition*, **21**, 409-421.

Perry, C., Ziegler, J. C., & Zorzi, M. (2007). Nested incremental modeling in the development of computational theories: The CDP+ model of reading aloud. *Psychological Review*, **114**, 273-315.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. E. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review*, **103**, 56-115.

Plourde, C., & Besner, D. (1997). On the locus of the word frequency effect in visual word recognition. *Canadian Journal of Experimental Psychology*, **51**, 181-194.

Rastle, K., & Coltheart, M. (1998). Whammies and double whammies: The effect of length on nonword reading. *Psychonomic Bulletin & Review*, **5** (2), 277-282.

Rastle, K., & Coltheart, M. (1999). Serial and strategic effects in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance*, **25**, 482-503.

Rastle, K., Harrington, J., & Coltheart, M. (2002). 358,534 nonwords: The ARC Nonword Database. *Quarterly Journal of Experimental Psychology*, **55A**, 1339-1362.

Rayner, K. (1975). Parafoveal identification during a fixation in reading. *Acta Psychologia*, **4**, 271-282.

Reicher G.M., (1969). Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology*, **81**, 274–280.

Reynolds, M., & Besner, D. (2002). Neighbourhood density effects in reading aloud: New insights from simulations with the DRC model. *Canadian Journal of Experimental Psychology*, **56**, 310-318.

Reynolds, M., & Besner, D. (2004). Neighbourhood density, word frequency, and spelling-sound regularity effects in naming: Similarities and differences between skilled readers and the dual route cascaded computational model. *Canadian Journal of Experimental Psychology*, **58**, 13-31.

Reynolds, M., & Besner, D. (2005). Basic processes in reading: A critical review of pseudohomophone effects in reading aloud and a new computational account. *Psychonomic Bulletin and Review*, **12**, 622-646.

Reynolds, M., & Besner, D. (2006). Reading aloud is not automatic: processing capacity is required to generate a phonological code from print. *Journal of Experimental Psychology: Human Perception and Performance,* **32** (6), 1303-1323.

Reynolds, M., Mulatti, C., & Besner, D. (2006). When looking dense is better than sounding dense: A dissociation between orthographic and phonological neighborhood density effects in reading aloud. Unpublished manuscript.

Roberts, M. A., Rastle, K., Coltheart M., & Besner, D. (2003). When parallel processing in visual word recognition is not enough: New evidence form naming. *Psychonomic Bulletin & Review*, **10**, 405-414.

Rosson, M. B. (1983). From SOFA to LOUCH: Lexical contribution to pseudoword pronunciation. *Memory and Cognition*, **11**, 152-160.

Rumelhart, D. E. (1970). A multicomponent theory of the perception of briefly exposed visual displays. *Journal of Mathematical Psychology*, **7**, 191-218.

Rumelhart, D. E., & Siple, P. (1974). The process of recognizing tachistoscopically presented words. *Psychological Review*, **81**, 99-118.

Scarborough, D. L., Cortese, C., & Scarborough, H. S. (1997). Frequency and repetition effects in lexical memory. *Journal of Experimental Psychology: Human Perception and Performance*, **3**, 1-17.
Scheerer, E. (1987). Visual word recognition in German. In A. Allport, D. MacKay, W. Prinz, & E. Scheerer (Eds.), *Language perception and production: Shared mechanism in listening, speaking, reading and writing* (pp. 227-244). London: Academic Press.
Sears, C. R., Hino, Y., & Lupker, S. J. (1995). Neighborhood size and neighborhood frequency effect in visual word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, **21**, 876-900.
Seidenberg, M. S., & McClelland, J. L. (1989). A distributed developmental model of word recognition and naming. *Psychological Review*, **96**, 523-568.
Seidenberg, M. S., Peterson, A., MacDonald, M. C., & Plaut, D. C. (1996). Pseudohomophone effects and models of word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **22**, 48-62.
Seidenberg, M. S., Waters, G. S., Barnes, M. A., & Tanenhaus, M. K. (1984). When does irregular spelling or pronunciation influence word recognition? *Journal of Verbal Learning and Verbal Behavior*, **23**, 383-404.
Selfridge, O. G. (1959). Pandemonium: a paradigm for learning. In D. V. Blake & A. M. Uttley (Eds.), *Proceedings of the Symmposium on Mechanisation of Thought Processes* (pp. 511-529). H. M. Stationary Office.
Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. *Acta Psychologica*, **30**, 276-315.
Sternberg, S. (1998). Discovering mental processing stages: The method of additive factors. In D. Scarborough & S. Sternberg (Eds.), *An invitation to cognitive science: Vol. 4. Methods, models, and conceptual issue* (pp. 703-863). Cambridge, MA:MIT Press.
Stolz, J. A., & Neely, J. (1995). When target degradation does and does not enhance semantic context effects in word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **21**, 596-611.
Strayer, D. L., & Kramer, A. F. (1994). Strategies and automaticity: I. Basic findings and conceptual framework. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **20**, 318-341.
Tabossi, P., & Laghi, L. (1992). Semantic priming in the proninciation of words in two writing systems: Italian and English. *Memory & Cognition*, **20**, 303-313.
Taft, M. & Russell, B. (1992). Pseudohomophone naming and the word frequency effect. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, **45A**, 51-71.
Taraban, R., & McClelland, J. L. (1987). Conspiracy effects in word pronunciation. *Journal of Memory and Language*, **26**, 608-631.
Taylor, T. E., & Lupker, S. J. (2001). Sequential effects in naming: A time-criterion account. *Journal of Experimental Psychology: Learning, Memory and Cognition*, **27**, 117-138.
Townsend, J.T. (1971). Theoretical analysis of an alphabetical confusion matrix. *Perception and Psychophysics*, **9**, 40-50.
Van Der Heijden, A.H.C., Malhas, M.S.M., & Van Den Roovaart, B.P. (1984). An empirical interletter confusion matrix from continuous-line capitals. *Perception and Psychophysics*, **35**, 85-88.
Van Selst, M., & Jolicoeur, P. (1994). A solution to the effect of sample size on outlier elimination. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, **47**, 631-650.
Visser, T. A., & Besner, D. (2001). On the dominance of whole-word knowledge in reading aloud. *Pychonomic Bulletin and Review*, **8**, 560-567.
Weekes, B. (1997). Differential effects of number of letters on word and nonword naming latency. *The Quarterly Journal of Experimental Psychology*, **50A**, 439-456.

Welford, A. T. (1952). The “psychological refractory period” and the timing of high-speed performance – a review and a theory. *British Journal of Psychology*, **43**, 2-19.
Whitney, C. S. (2001). How the brain encodes the order of letters in a printed word: The SERIOL model and selective literature review. *Psychonomic Bulletin & Review*, **8** (2), 221-243.
Whitney, C. S., Berndt, R. S., & Reggia, J. A. (1996). Simulation of neurogenic reading disorders with a dual-route connectionist model. In J. A. Reggia, E. Ruppin, & R. S. Berndt (Eds.), *Neural modeling of brain and cognitive disorders* (pp. 201-228). Singapore: World Scientific.
Whitney, C. S., & Lavidor, M. (2004). Why nonword length only matters in the left visual field. *Neuropsychologia*, **42**, 1680-1688.
Winnick, W. A., & Daniel, S. A. (1970). Two kinds of response priming in tachistoscopic recognition. *Journal of Experimental Psychology*, **84**, 74-81.
Yap, M. J., & Balota, D.A. (2007). Additive and interactive effects on response time distributions in visual word recognition. *Journal of Experimental Psychology: Learning Memory and Cognition*, **33**, 274-296.
Yap, M. J., Balota, D. A., Tse, C., & Besner, D. (2008). On the additive effects of stimulus quality and word frequency on lexical decision: Evidence for opposing interactive influences revealed by RT distributional analyses. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **34**, 495-513.
Young, A. W., & Ellis, A. W. (1985). Different methods of lexical access for words presented to the left and right visual hemifields. *Brain and Language*, **24**, 326-358.
Zevin, J. D., & Balota, D. A. (2000). Priming and attentional control of lexical and sublexical pathways during naming. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **26**, 121-135.
Ziegler, J. C., Perry, C., & Coltheart, M. (2000). The DRC model of visual word recognition and reading aloud: An extension to German. *European Journal of Cognitive Psychology*, **12**, 413-430.

APPENDIX

A. Nonword set Chapter 2

One or more orthographic neighbours				Zero orthographic neighbours	
Short		Long		Short	Long
nonword	***baseword***	**nonword**	***baseword***	**nonword**	**nonword**
beceo	*becco*	buocche	*brocche*	bluva	bruchio
bluco	*bruco*	breglia	*briglia*	bovre	braighe
burpo	*burro*	bleccia	*breccia*	bunpo	bleuche
cavua	*cavia*	chiozza	*chiazza*	bupio	buospio
cedlo	*cedro*	cenghia	*cinghia*	buppi	biucche
ceflo	*ceffo*	cioccio	*ciuccio*	crasa	chirria
catra	*cetra*	creccio	*cruccio*	cufli	chiuspo
cluva	*clava*	chiusco	*chiosco*	cusbo	crausco
cubra	*cobra*	cheatta	*chiatta*	dripo	drussio
felfa	*felpa*	friccia	*freccia*	fafre	friofro
fiulo	*fiala*	foschio	*fischio*	fluco	freusio
fraca	*frana*	frantia	*frangia*	gedre	geresco
gemua	*gemma*	giungra	*giungla*	geflo	gelagri
gerfa	*gerla*	giustra	*giostra*	gnasa	gnattro
ghito	*ghiro*	ghionda	*ghianda*	groze	grompio
gnolo	*gnomo*	gnaffio	*graffio*	liplo	luorlio
gurme	*germe*	gruglia	*griglia*	lumio	liospuo
melfa	*melma*	manvria	*mandria*	luofa	luostia
milpa	*milza*	muccheo	*mucchio*	meluo	miorfio
mucua	*mucca*	muscheo	*muschio*	nurio	neucche
murlo	*merlo*	mucchia	*macchia*	nuzao	nerghio
piuba	*piuma*	piccheo	*picchio*	pivvo	praschi
piuba	*piuma*	piustra	*piastra*	quoso	quochia
siupe	*siepe*	strullo	*strillo*	sfepo	sproghi
sludo	*scudo*	spiuzzo	*spruzzo*	soplo	strisso
sviva	*stiva*	svrazzo	*sprazzo*	tiafi	trippio
tipre	*tigre*	truglia	*triglia*	vreva	vregghi
vesba	*vespa*	voschio	*vischio*	zatro	zianglo

B. Nonword set Chapter 3

List A		List B	
Low TLC	High TLV	Low TLC	High TLV
DAVIUTA	DEBEFIO	DUCAZIA	DOBERSO
DILICUA	DISULMA	DUTILIA	DROSORO
DIVILIA	DEMEBRO	DUZIALA	DABENEO
DIZILVA	DREBENO	DILCAVA	DEBREMO
FAVULIA	FEBERMO	FOSPATA	FEBARIA
FIVATUA	FEBLECA	FULIVUA	FREBANO
FILICUA	FRESUBO	FURZICA	FOBARGO
GACUVIA	GABENIA	GARITIA	GORIBIO
GALALIA	GRONERO	GULATIA	GOBERSO
GAVACUA	GOBREMO	GATICUA	GOMEBEO
OCAVITA	OBEMINO	OPICUVA	OMEGANO
OCILIVA	OBIMATO	OTIRIVA	OMESIBA
OCIVATA	OFORIBO	OVICUCI	OPOBINA
OLILICI	OLEBERA	OVATULA	OBEREFO
PALIVUA	PURIBIO	PLIVAVI	PORUTIA
PLATIVA	PRISICO	PATUZIA	PEBERIO
RAVICUA	REBRESO	RILIZIA	REMERMO
SANELIA	SANEBRO	SICULIA	STIPOBA
SACUVIA	SCUBELA	STICUCA	SEMOFRO
SUCIZIA	SEBERNO	SAZILIA	SEBREMO
SLUCAVA	SBEMERO	SCUVALA	SOMEBIE
TOLTOVA	TERADIO	TUCAZIA	TREMEFO
TRISIPA	TRESENO	TICUVIA	TINAIGO
VILACUA	VEBRESO	VUCITRA	VENORSO
VOZILIA	VREBEMO	VAVILIA	VEMERSO

C. Nonword set Chapter 4

TLC balanced		TLC not balanced	
Short	Long	Short	Long
DEFEO	DIVILIA	DUCAO	DISULMA
DROME	DUZIALA	DRIPO	DEBEFIO
FEQUE	FLICUAI	FAFRE	FREUSIO
FODRO	FOSPATA	FLUCO	FEBARIA
FOQUE	FURZICA	FLUPA	FEBLECA
GEDRE	GIARTIA	GEFLO	GIORBIO
GHEBE	GACUVIA	GHEIA	GABENIA
GHEMO	GHIVILA	GHEDO	GHIRAMA
GNEBE	GNALPIA	GNEVA	GNOMPIO
GNESO	GNAUCIA	GNALA	GNEBRUO
OBBEO	OLILICI	OPPUA	OMESBAI
OBOBE	OTRIVIA	OCACE	OFROBIO
OBROE	OPICUVA	ORCAO	OLEBREA
ONEBO	OCILIVA	OCITA	OBEMINO
ONENA	OLAGHIA	ODADO	OPROGHI
ORIRO	OLICHEA	ODALO	ONERCHI
OSEBO	OCAVITA	OZICA	OPOBINA
OVEGO	OCIVATA	OTIDO	OMEGANO
OZEMO	OVICUCI	OZICA	OBIMATO
PEMIO	PLIATIA	PUFLI	PREISCO
PEQUO	PLIVAVI	PILVO	PUORTIA
POBEO	PLAUVIA	POCAO	PIURBIO
SEFLO	SUCIZIA	SOFLA	SANEREO
SESME	SANELIA	SUSIO	SOMEBIE
SIOFO	SAVICUA	SOPLO	SEBENEO
SMEBO	SCULVIA	SUMIO	SCUBLEA
SVABE	STICUCA	SFEPO	STIPOBA
TEOBA	TUCAZIA	TIAFI	TINAGIO
TIBBO	TUOGLIA	TIMMA	TREGLIO
TREBE	TRISPIA	TISBO	TEARDIO

D. Nonword set Chapter 5

List A		List B	
Many-N	Zero-N	Many-N	Zero-N
BLAVE	BLIRP	BLICK	BLICH
CHACK	CHYTH	CHARP	CHYLE
CLECK	CRAUN	CLOSS	CLIGH
CREAT	CRIRR	CRAFE	CRYLE
DATCH	DRICH	DRAVE	DRURP
FRICK	FLIFE	FLATE	FLAUB
GAINT	GHAIF	GATCH	GHISC
GIGHT	GLITH	GLEAT	GLYME
GLAVE	GLIEK	GRINT	GRULE
PETCH	PLAFF	PAUNT	PLECH
PLAME	PLARM	PLARE	PLAUB
PRITE	PLOAM	POUTH	PLIVE
POTCH	PLUTT	PRIPE	PLYTH
PRIVE	PRAIF	PLINK	PRAUT
SLIPE	SKAUK	SAUNT	SLIEN
SCORT	SCROU	SCALL	SLIER
SPOOP	SLIRM	SLINE	SPAPH
SLOUT	SPYTH	SPOOT	SPLEA
TATCH	TRARC	TOUTH	TRURL
TRIVE	TWAUL	TRAVE	TROAR
BINCH	BLIFE	BLANE	BLERF
CHASS	CHERF	CHONE	CHYBE
CRAME	CRYBE	CRASE	CRERF
CROSE	CRARN	CREET	COOSH
DORSE	DREWT	DREAK	DRIRR
FRESS	FLEBB	FOUSE	FLENE
GOUSE	GHIRM	GRABE	GHYTH
GRAGE	GRERG	GRARE	GRIGH
GRAME	GRURK	GRASH	GRERF
PENCH	PLEEM	POOSE	PRAUB
PORSE	PLERB	POUSE	PLEWN
PROME	PRERG	PROSS	PRECH
PRINE	PRETE	PROPE	PREUM
PROWN	PROCH	PRIBE	PRURB
STAFE	SKOAM	SEAVE	SKURR
SPAME	SLEFF	SMORE	SKASS
STOOK	SLERG	STABE	SLESE
STASE	SLUBE	STORN	SLOMB
TORSE	TEECE	TOUSE	TRURR
TRUSH	TWOAR	TRIME	TROAM

E. Word and nonword sets Chapter 6

Word		Nonword	
Low TLC	**High TLC**	**Low TLC**	**High TLC**
BALIA	BENDA	BILZO	BERCA
CELLA	CENNO	CITRA	CESBO
CLAVA	CREMA	CLARO	CRENE
COCCO	COSMO	CALSA	CABRO
COLPA	COBRA	COLVO	CONFE
DANZA	DENTE	DILCE	DERDO
DITTA	DOSSO	DIACO	DREGO
FALCO	FORNO	FILPA	FENTE
FALDA	FARRO	FLOCA	FRONA
FIALA	FIENO	FIAVA	FIEBA
FIATO	FIORE	FILCO	FIRBA
FOLLA	FONTE	FUTTA	FOSBA
GARZA	GAMBO	GUTTO	GUMMA
GAZZA	GONNA	GOLCO	GREDO
GIOIA	GEMMA	GIUCO	GEBIO
LACCA	LEMBO	LADLO	LEBRE
MAZZA	MAMMA	MAPIA	MEFFA
MOLLA	MERLO	MITUA	MIEBE
MUCCA	MANGO	MILGA	MANBA
MULTA	MENTA	MALCA	MENZO
PACCO	PENNA	PINIA	PONNA
PALCO	PERNO	PAUCO	PERTO
PANCA	PRETE	PISCA	PONBE
PULCE	PLEBE	PALDA	PONNO
RAZZA	RENNA	RATIO	RENTA
RICCO	ROSPO	RUCCA	RASSA
RULLO	ROMBO	RULTA	RISMO
SALTO	SASSO	SALCA	SEBIA
STIVA	SIEPE	SAITA	SUOMA
SUOLA	SUONO	SCAPA	SMAGO
TACCO	TONNO	TITTO	TARSO
TALPA	TORTO	TALGA	TARBA
TAPPO	TOSSE	TALTO	TONGO
TARLO	TORBA	TROLA	TREVE
TASCA	TRENO	TOITA	TIBRE
TEDIO	TONFO	TUFLO	TUOMO
TAZZA	TORRE	TITIA	TENBA
VALLE	VERME	VITTA	VEMBO
VASCA	VESTE	VALTO	VESBA
VILLA	VENTO	VILPE	VEBRO

F. Word and nonword sets Chapter 7

WORD	NONWORD (baseword)	WORD	NONWORD (baseword)
BADGE	BAIME (baize)	LEASH	LERGE (ledge)
BEACH	BEART (beard)	LUNCH	LAMPH (lymph)
BATCH	BENTH (bench)	MIRTH	MORSH (marsh)
BEAST	BERCH (beech)	MOOSE	MOURD (mound)
BLAZE	BLAVE (blade)	NERVE	NARSE (nurse)
BLISS	BROCK (block)	NIECE	NATCH (notch)
BLOOM	BLEAM (bream)	PEACE	POTCH (patch)
BRIDE	BRACK (brick)	PEACH	PHOSE (phone)
BRINK	BROOL (broom)	PIECE	PUTCH (pitch)
CHAIR	CHERT (chest)	PLANE	PLONT (plant)
CHURN	CHERK (cheek)	PLUME	POUTH (pouch)
CHICK	CHISS (chess)	PRAWN	PRASK (prank)
CLOAK	CLATH (cloth)	PRIZE	PRIVE (pride)
CLOCK	CRINE (crime)	PLANK	PLAVE (plate)
CLEFT	CLOID (cloud)	QUEEN	QUERT (quest)
CLOWN	CLOME (clove)	QUILL	QUIRT (quilt)
CREED	CLEAM (cream)	RANCH	REAGN (reign)
COUCH	CANCH (conch)	RHYME	ROACK (roach)
COAST	CRUCK (crack)	SAUCE	SCASE (scare)
CRASH	CREEB (creek)	SCARF	SCODE (score)
CROWD	CRAWN (crown)	SCOOP	SCOKE (scope)
DREAM	DRISS (dress)	SCOUT	SCREP (scrap)
FARCE	FLITH (faith)	SHADE	SHASE (shame)
FIRTH	FROOR (floor)	SHANK	SHORK (shark)
FIGHT	FLUVE (flute)	SHAPE	SHEST (sheet)
FLAKE	FLAVE (flame)	SHAWL	SHECK (shack)
FLAIR	FLINK (flank)	SHEAF	SHERN (sheen)
FLINT	FLOKE (fluke)	SHORE	SHICK (shock)
FOUNT	FORVE (force)	SHIRT	SHELT (shelf)
FRAME	FROUD (fraud)	SHRUB	SHOST (shout)
GLOBE	GLASE (glade)	SKILL	SLIRT (skirt)
GLOSS	GLOUM (gloom)	SLASH	SLONG (slang)
GOOSE	GRAFE (grape)	SNACK	SNACE (snake)
GRAPH	GLEED (greed)	SPORT	SPOUN (spoon)
GUEST	GRODE (grove)	SPIKE	SPIME (spice)
GUILT	GRILE (guile)	SPATE	SNAVE (snare)
HORSE	HEASH (heath)	TRICE	TRIMP (trump)
HOUND	HANCH (hunch)	TRAIN	TRUCH (truck)
JAUNT	JAICE (juice)	VALVE	VARGE (verge)
LARCH	LOTCH (latch)	WITCH	WHAME (whale)