# Discovering large grain-sizes in a transparent orthography: insights from a connectionist model of reading for Italian

Giovanni Pagliuca

Department of Psychology, University of York, York

and

Padraic Monaghan

Department of Psychology, Lancaster University, Lancaster

3 April 2009

Corresponding author:

Giovanni Pagliuca

Department of Psychology, University of York

York YO10 5DD, UK

Email: g.pagliuca@psych.york.ac.uk

Tel: +44 (0)1904 432885

word count: 5186

Abstract

Classic connectionist models of reading have traditionally focused on English, a language with a quasi-regular (deep) relationship between orthography and phonology, and very little work has been conducted on more transparent (shallow) orthographies. This paper introduces a parallel distributed processing (PDP) model of reading for Italian. The model was explicitly developed in order to deal with polysyllabic words and stress assignment. One of the core issues regarding such class of models is whether they can show sensitivity to large grain-sizes, as documented by the existence of morphological and neighborhood effects in nonword reading aloud showed by native Italian speakers (Burani, Marcolini, De Luca, & Zoccolotti, 2008; Arduino & Burani, 2004). The model is successful in simulating such sensitivity, previously accounted for by dual route architectures, and is also tested in order to simulate stress consistency effects. The model provides clear evidence that large grain-sizes in the orthography to phonology mapping can be discovered even in a model trained with almost perfectly shallow stimuli.

*Keywords*: shallow orthography; modeling; naming; morphology; grain-size

## Introduction

Connectionist models of reading have been originally developed to explore the general cognitive architecture of the reading system but also the specific psycholinguistic effects that have been documented over the years for the English language. Very little work has been conducted to extend the parallel distributed approach or other modeling approaches to reading beyond the English language, despite the implicit claim that the principles that govern the reading system are universal and should therefore apply to all orthographies, both alphabetic and logographic, deep and shallow. PDP models of reading for English have been trained on small sets of monosyllabic words and have simulated a vast collection of behavioural effects related to monosyllabic single word reading (Plaut, McClelland, Seidenberg, & Patterson, 1996; Harm & Seidenberg, 1999, 2004). One of the main practical difficulties in exporting a PDP architecture developed for English to other orthographies is due to potential distinctions in terms of the syllabic properties of the language. For example, some languages have very few monosyllabic words, and therefore a monosyllabic model would not be representative of the language as a whole. Or alternatively, the modeller must account for the constraints that a polysyllabic structure would impose on the model, such as those posed by stress assignment. A few attempts have nonetheless been made to model psycholinguistic effects in languages such as German and French within a general PDP framework.

Hutzler, Ziegler, Perry, Wimmer, and Zorzi (2004) adapted Plaut and collaborators' (1996) feedforward network to read German monosyllabic words, and found a general advantage of this model compared to the English version in speed of learning, despite the numerous similarities of the two languages in terms of orthographic and phonological complexity (but not in the mapping between the two, German being more regular than English). Ans, Carbonnel and Valdois (1998) trained a polysyllabic connectionist network to read French. The model was trained on a large corpus of mono and polysyllabic French words, and could read successfully 96.32% of them, and account for accurate nonword reading, frequency and consistency effects, and simulate phonological dyslexia as well.

The reading process has generally been defined in the modeling literature in terms of forming a mapping between visual symbols and phonemes or syllables (Plaut et al., 1996), and as a mapping between visual information and meaning (Harm & Seidenberg, 2004). The orthography to phonology pathway and the orthography to semantics pathway are distinct in terms of the nature of the mapping between the representations. Spelling-sound correspondences in languages with alphabetic writing systems (such as English and Italian) have two critical properties. They are *systematic*, in that words that are written similarly have similar pronunciations. They are also *componential*, with individual letters, or pairs of letters, within the written word corresponding to certain phonemes in the pronunciation. In English, the mapping between orthography and phonology has many exceptions for irregular words (Plaut et al., 1996), though the properties of systematicity and compositionality remain generally the case. In contrast, mappings between written words and their meaning are largely *arbitrary* (Monaghan & Christiansen, 2006; Saussure, 1916) and non-compositional. Words that are written similarly are likely to have very different meanings, and, equally, individual letters within the word do not provide information about meaning. Reading for pronunciation, then, entails a system that responds to the pronunciation of individual letters, or small groups of letters, within the word (Ziegler & Goswami, 2005), whereas reading for meaning requires a system that processes the word in its entirety.

Learning to read can therefore be described as a process of learning to find shared "grain sizes" (or "what maps into what" pairs) between orthography and phonology (Ziegler and Goswami, 2005). According to this theoretical framework, these grain sizes are language specific and allow for an efficient mapping between the two levels or representation. In this view the ability to identify and make optimal use of orthographic clusters in reading plays a major role in defining the ability to read. This identification process is dependent on the characteristics of the spelling to sound mapping and its consistency, which varies greatly across different languages. In some orthographies, one letter can have more than one pronunciation, as is the case for English and Danish, whereas in others it is nearly always pronounced in the same way, as in Italian or Greek. In the same way, in some orthographies, a phoneme can have multiple spellings (e.g., English), whilst in others generally only one (e.g., Italian). Italian and Greek, then, have a high degree of spelling-to-sound consistency, whereas English and Danish have a high degree of spelling inconsistency. These differences are reflected in the reading development of children exposed to the two types of orthographies (highly consistent versus highly inconsistent): children learning shallow orthographies show a rapid reading acquisition compared to children learning deep orthographies (Seymour, Aro, & Erskine, 2003). For example, at the end of Grade 1 Greek children can read 90% of familiar words correctly, while the level of accuracy is only 34% for Scottish English children, with Danish being somewhat intermediate, with 74% correct reading. These percentages show that there is a clear correlation between level of performance and consistency in the spelling to sound mapping, with highly consistent orthographies being learned faster than inconsistent ones.

The spelling inconsistency is related to the problem of orthographic *granularity*: relying on smaller grain sizes (e.g., single letters) in deep orthographies would result in a high degree of irregularity in the orthography-phonology mapping, and so larger grain sizes (e.g., bigrams or trigrams) are generally more consistent. Readers have to learn to rely on different grain sizes according to the level of inconsistency present in the orthography to phonology mapping, and the size of the orthographic window in recoding might vary according to the specific language considered. According to the psycholinguistic grain size theory (Ziegler & Goswami, 2005), readers of shallow orthographies will tend to rely more on small grain sizes, like letters, as they prove consistent and also highly frequent units within the language; conversely, readers of deep orthographies will develop different recoding strategies at more than one grain size, varying the window of the orthographic recoding. However, the reliance on small grain sizes does not automatically exclude the possibility that large grain sizes might help reading development in young readers and define reading strategies in skilled adult readers even for languages with transparent orthographies.

Models which implement distributed forms of representations can successfully discover these different functional units, as shown by Pagliuca and Monaghan (2008). Pagliuca and Monaghan (2008) lesioned an adapted version of Harm and Seidenberg's (1999) model of reading and showed that the damaged network could perform better when tested with words containing multiletter graphemes (as SH in the word *shine*) as compared to control words with no multiletter graphemes. The model could reproduce the multiletter graphemes at the phonological level because it had learned to associate the two letters belonging to the grapheme to one single phoneme, and could generate the correct target even when activation from each single letter slot was reduced, as the combined activation of the pair of letters was still enough to generate the right phoneme. As the model did not use localist units in the orthographic layer to code for multiletter graphemes, but used only single letters, this sensitivity shows that parallel distributed neural networks can successfully detect shared grain sizes in the orthography to phonology mapping, at least at the level of two-letter graphemes.

Traditionally, PDP models have been successfully implemented to explore consistency effects in word and nonword reading, effects that require the model to capture different shared grain sizes in the orthography to phonology mapping (Zevin & Seidenberg, 2006). Given the potential that PDP networks have to discover appropriate grain sizes in the orthography to phonology mapping, a core question is whether such a class of models can discover grain sizes larger than a single unit in an almost completely regular and shallow orthography, one for which the single-letter to single-phoneme mapping alone allows for an almost perfect compositional recoding of orthography, and therefore almost perfect pronunciation. A way to test this possibility is offered by the Italian language, whose spelling to sound mapping is almost entirely regular, with a one to one correspondence between letters and phonemes for most letters. Nonetheless, several effects at the lexical and morpholexical level have been documented for Italian, suggesting that Italian readers show sensitivity to reading units larger than the single letter or pairs of letters in word naming and word recognition.

In this paper we report a fully distributed model of reading for Italian, conceived in order to expand the PDP framework to transparent orthographies and further investigate the relationship between orthography and phonology in the light of a grain size perspective on reading. We first report the orthographic properties of Italian and studies that indicate a larger grain-size than the single letter in this language.

## Properties of Italian orthography and phonology

Italian is an alphabetic orthography with an almost entirely compositional one-to-one mapping between spelling and sound. In Italian each letter regularly translates in a single phoneme, with few exceptions fully predictable by the orthographic context. For example, the letter B is always pronounced /b/, irrespective of the surrounding letters, but the letters C and G can obtain two different pronunciations according to the following vowel: /ʧ/ and /ʤ/ when followed by the vowels I and E, /k/ and /ɡ/ when followed by the vowels O, U, A or by the letter H, which in Italian is unvoiced. The letter G can also be pronounced as a liquid if followed by the letter L in combination with I, as /ʎ/ (see Burani, Barca & Ellis, 2006, for a full list of contextual rules for Italian).

There are very few monosyllabic words in Italian, and most of them are function words. Given the polysyllabic structure of most words, Italian readers are confronted with the problem of stress assignment, which is perhaps one of the most interesting feature of Italian. Stress assignment in

Italian is considered quasi-regular: for the vast majority of words (about 80%), stress is placed on the penultimate syllable, as in *al'bergo* (hotel), but there are many exceptions to this rule, with stress placed on the antepenultimate syllable in 18% of cases, as in '*albero* (tree) (see Thornton, Iacobini, and Burani, 1997, for the estimated count). A small proportion of words, about 2%, have final stress, but in these cases the words are marked with diacritics in the written form, as in *papà* (dad), therefore final stress assignment is fully predictable from these orthographic cues. Words with the most frequent (dominat) stress pattern are considered to be "regular", while words with the less frequent one (non-dominant) are considered "irregular".

Colombo (1992) was the first author to explore systematically stress assignment in Italian, and found an interaction between regularity and frequency, with regularly stressed low frequency words read aloud faster then irregularly stress low frequency words. This effect has been interpreted within a dual route framework (such as Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001): according to this interpretation, low frequency words are more likely to be processed by a non-lexical mechanism, which generates a "default" regular stress pattern. This pattern is in conflict with the correct "irregular" stress assignment activated in the lexicon, and causes a delay in pronunciation.

However, stress assignment in Italian words also correlates with the word ending (nucleus of the penultimate plus the final syllable): for example, words ending in *–oro* are mainly regularly stressed , with only about 17% of these words being irregularly stressed, while words ending in –*ola* take mostly irregular stress, with only a minority of them being stressed regularly. Four different combinations can result from this classification: regularly stressed words with many friends, regularly stressed words with many enemies, irregularly stressed words with many friends, irregularly stressed words with many enemies (Burani & Arduino, 2004).

Colombo (1992) found that low frequency irregularly stressed words with a high number of stress friends were named aloud faster than low frequency irregularly stressed ones, but no such effect of stress neighbourhood was found for regularly stressed words. This finding supported the claim that stress assignment in regular words is generated by default.

However, Burani and Arduino (2004) found instead that the effect of stress neighborhood was not simply restricted to irregular words, but extended to regular words as well. Moreover, no effect of stress regularity in reading aloud low frequency three and four syllables words was found. The authors suggested an interpretation of the finding that does not involve a system of default rules, but is in line with general parallel distributed principles: the extraction of stress assignment cues (word endings) is favoured when these cues are shared by many words, and is not favoured when shared by only a few. This debate about stress regularity or stress consistency would benefit greatly from a model that could simulate stress assignment for Italian words, which is one of the aims of this paper.

The Italian spelling to sound mapping has been studied extensively in the last few decades and a few benchmarks have been established for this transparent orthography. The strong regularity in the spelling to sound mapping might alone promote the use of a nonlexical reading strategy with word naming primarily mediated by a sublexical code, and reliance on the use of grapheme-to-phoneme correspondence rules, as suggested in previous studies (Frost, Katz, & Bentin, 1987). However, a marked lexicality effect and a frequency effect have been documented for Italian, even when using completely transparent stimuli (Pagliuca, Arduino, Barca, & Burani, 2008), effects that cannot be explained solely by the use of sublexical conversion mechanisms at the level of single letters or bigrams.

Lexical contributions have nonetheless been found in nonword reading as well, challenging the claim that nonwords are solely read via a nonlexical serial mechanism (Frost et al., 1987), suggesting again that Italian readers do not simply rely on a set of rules to convert single letters onto phonological representations when reading nonwords. Arduino and Burani (2004) found that nonwords which had a large cohort of lexical neighbours (nonwords that vary from other words by one letter only) were named faster than nonwords which had very few neighbours. This effect was found irrespective of the frequency of the neighbouring words. The effect was ascribed by the authors to the contribution of a lexical lookup mechanism alongside a grapheme to phoneme set of rules within a dual route framework (Coltheart et al., 2001) with both mechanisms being active when reading nonwords as words.

What these results primarily show is that Italian readers can discover and make use of multi-letter representations perhaps up to the word level shared by many words which provide information that goes beyond the reach of a strictly rule-base mechanism.

Morphological effects have also been documented for Italian. Italian readers seem to benefit from the presence of a morpheme in reading nonwords (Burani, Marcolini, De Luca, & Zoccolotti, 2008). Nonwords containing real morphemic units (*donnista*, made up of a real root *donn-* and a real suffix *–ista*) were named faster than control nonwords (*dennosto*, which has no real root and no real suffix) matched for bigram frequency and length both by adult readers, young readers (sixth grade) and dyslexic children (Burani et al., 2008). Dyslexics and younger children (second and third grades) also benefited from morphological structure in reading aloud real words. The authors suggest that the morpheme is an effective reading unit for Italian, complementary to whole-word lexical information, which again represents a unit larger than single letters or pairs of letters. It may well be that the morpheme is one of the units employed by the reader of Italian in decoding the orthography into a phonological form. Morphemes occur highly frequently in the Italian lexicon, and the morphological effect could therefore be an effect due to processing of high-frequency trigrams, or higher order n-grams that assist in the mapping between orthography and phonology. We assume that any regularity between letters and sounds that frequently occurs is likely to be exploited by the reader. In line with this view the morphological effect described in this study is likely to stem purely from orthographic and phonological redundancy and it is possible that it emerges within the orthography to phonology pathway rather than the orthography to semantics pathway. In Italian, a morphologically rich language, there is evidence that morphological effects in word naming are non-semantic in nature (Burani, Dovetto, Spuntarelli, Thronton, 1999), contrary to morphological effects reported for English (Plaut & Gonnerman, 2000). In the study by Burani and collaborators (1999), the authors found that morphologically complex nonwords which had a high degree of semantic interpretability were recognized faster than nonwords with a low degree of semantic interpretability in a lexical decision task, but that was not the case in a naming task. In naming nonwords aloud, there was no effect of their degree of semantic interpretability. Yet, nonwords containing real morphemes were still named faster than nonwords which did not include any morpheme, confirming the presence of a general advantage for morphologically complex stimuli. The authors suggest that the effect might depend on a lexical (morphological) non-semantic pathway rather than on a lexical semantic one.

In this respect, the model we introduce here is suitable for testing the hypothesis that in Italian morphological effects in naming aloud do not necessarily depend on the semantic pathway, as our model does not employ a semantic layer nor an orthography to semantic pathway. The model has then to develop internal representations for morphemes during learning the mapping between orthography and phonology only. Though Burani et al. (2008) controlled for bigram frequency, they did not (and presumably could not) control for these higher-order n-grams that are present in the language and so it remains a possibility that the morpheme effects are due to the multiple grain-sizes useful in the orthography-phonology mapping rather than requiring an intermediate morphological route. One of the aims of the Italian model was to establish whether such morphological effects were consistent with a model that can only compute the multiple regularities between written and spoken words, and that does not contain a semantic or morphological system.

It seems evident from these reported studies that Italian readers can exploit the spelling to sound mapping beyond the smallest possible grain size (single letter), even when the mapping itself allows for an apparently sufficient and efficient one-to-one letter to phoneme recoding strategy. Italian readers show sensitivity to different grain sizes (graphemic, morphological, lexical), according to the stimuli they are asked to name aloud. However, this sensitivity does not necessarily imply that graphemic, morphemic and lexical information is stored and accessed independently, nor does it entail that there are intermediate units of representation between orthography and phonology that code explicitly for this kind of information. Nor does it require separate mechanisms that interact to generate the observed effects. A single-route PDP model that maps orthography onto phonology can in principle be used to explore some of the effects described for reading in Italian, and indicate how multiple grain-sizes can be discovered in learning to map orthography onto phonology in order to determine how such grain-sizes can explain the observed psycholinguistic effects of reading in Italian. More importantly, a PDP architecture could potentially discover the appropriate grain sizes that emerge in mapping Italian orthography onto phonology – we argue that such grain sizes are a matter of empirical discovery from the entire lexicon, and their role for particular words, or subsets of words, cannot be determined by examining the general properties of the Italian language. This result would be even more striking given the extreme regularity of the mapping.

Next, a PDP model of reading is described, which, in line with other fully distributed architectures, does not implement localist units to represent lexical information nor does it instantiate distinct lexical and sublexical mechanisms to generate the appropriate phonology for each word and to assign the correct stress pattern.

## Simulation1: Modelling Italian Reading

### *Method*

#### *Architecture and Representation*

The architecture of the model is closely based on Harm and Seidenberg's (1999) model of reading, and is shown in Figure 1. The orthographic layer comprised 476 units, the hidden layer had 100 units and the phonological layer contained 204 units. A set of 50 cleanup units was added to the network and connected bidirectionally to the phonological layer in order to create a set of phonological attractor units. The phonological layer was self-connected to itself with connection weight of 0.75. This was to ensure that the phonological attractor units were involved in maintaining the phonological representation at the output – due to these self-connections, the activitation would gradually decline, unless the units connecting to it provided a boost to the activity. Orthography in the model was represented in a slot based manner, with 3 slots for the onset, 2 for the vowels, and one for the coda for each syllable. The last syllable had no slot for the coda as typically Italian words do not end with a consonant (the few words that do end in a coda tend to be loan-words, and we omitted these from the lexicon). Up to three syllables could be represented in the orthographic layer, for a total of 17 slots (6 slots each for the first 2 syllables, 5 slots for the third syllable). The syllables in a word were left aligned, with monosyllabic words occupying up to the first 6 slots, two-syllables words occupying up to the first 12 slots, and three syllables words occupying all 17 slots. Within each letter position slot, a total of 28 distinct letters were represented in the model's input. Vowels with accents were represented in the orthography as distinct letters.

Phonology in the model was implemented in terms of phonological features, in line with recent PDP models of reading (Harm & Seidenberg, 1999, 2004). Each phoneme was described by a set of 11 standard binary phonological features (Harm & Seidenberg, 1999). An extra feature was added to the phonological features in order to distinguish stressed vowels from unstressed ones, bringing the total number of features to 12 for each phoneme. The set of phonological features used was taken from Canepari (1980). Open and closed vowels (i.e., open and closed "o" and open and closed "e") were treated as separate phonemes.

*Training corpus*

In order to create a sizeable corpus of words to train the model with, two different databases were combined. Orthographic forms were initially extracted from the "Corpus e Lessico di Frequenza dell'Italiano Scritto" (CoLFIS) database (Bertinetto, Burani, Laudanna, Marconi, Ratti, Rolando, & Thornton, 2005). This database contains frequency information from a corpus of 3 million words, and this frequency information was extracted as well. Plurals and inflected forms were included. The database contained both mono and polysyllabic words. Words beginning with the letters H, J, Y, W, X were excluded from the corpus, as in Italian they appear almost only in loan words. Words containing the letters J, Y, W, X in any other position in the word were excluded as well for the same reason. Two more types of information about the lexicon were needed: syllabic boundaries for each word and stress position. This information was extracted from the De Mauro Italian Dictionary (De Mauro, 2000). This database contains stress placement information and syllabic boundaries: each word is split into its constituent syllables (typographically separated by a hyphen) and the stressed syllable is marked by the use of a diacritic above the stressed vowel. Only primary stress is represented. The two databases were then co-indexed and only those words that were present in both dictionaries were used for training the model. A total of 29336 words resulted from this combination. Only monosyllabic, bisyllabic and trisyllabic words with three or fewer vowels in the nucleus were further selected, resulting in a total of 9911 words.

A phonological representation for each word was created using an algorithm to translate orthography onto phonology. Double consonants are true geminates in Italian and were coded as two separate phonemes, one each in the coda and onset of adjacent syllables. Diphtongs were not coded as different phonemes but were broken down into their constituent vowels. Frequency for each word was capped at 1000 and then compressed (square root compression).

*Training and testing*

The model was trained with the continuous recurrent backpropagation algorithm (Harm & Seidenberg, 2004). In the first stage of training, the phonological attractor system learned to represent all the words in the lexicon up to the point where the mean square error for the output of each pattern was below 0.01. For 4 time ticks all phonological units were clamped with the appropriate values for the target word. Then, for ticks 5-11 the output of each phonological unit was compared with the actual value of the word, and the difference was propagated backwards thought the network, generating error gradients for each word, and the weights were then updated according to the backpropagation learning algorithm to enable the model to produce an output closer to the target. The trained weights for the phonological attractor system were then fixed in the reading model.

In the second stage of training, the model was trained to map between the orthographic input of the model and the phonological output. A learning rate of 0.005 and momentum of 0.9 were used. The model was trained for 1.2 million word presentations, after which training was stopped and the model's performance was assessed.

*Results and Discussion*

Naming accuracy and sum squared error (SSE) were computed to test the model's general performance. Euclidian distances were computed for each phoneme and the closest phoneme to the target was selected and reported as the models' final solution. A word was judged to be generated correctly if all of its phonemes were reproduced in each slot. SSE was computed from the model's output, as well as determining the nearest phonological output target.

After 1.2 million word presentations the model could read correctly 93.7% of all words. Of the errors 10% were classified as true phonological errors (*debacle* reads as *debaple*, with a phoneme substitution), while 26% were classified as "stress placement" errors (*im'pala* read as *'impala*).

64% of errors affected the reproduction of the vowels “o” and “e” along the open/closed dimension (an open “o” generated instead of a closed “o”). However this last type of response is usually not classified as an error in the behavioural literature, due to large regional variation in the use of these 2 vowels. With the exclusion of this type of error, the model was 98% correct at reading the words in the corpus. The model was therefore successful in extending previous models of reading to apply to reading Italian. The model demonstrated that polysyllabic phonological output could be produced, and appropriate stress could also be determined for a large, representative lexicon of Italian.

*Nonword Reading*

An additional benchmark for computational models of reading is nonword performance. A model that fails to generalise to reading novel stimuli that conform to the general pattern of the language, as human reader’s are able to do, would not provide an effective model of human reading behaviour. Three sets of nonwords were selected from the literature: 48 bisyllabic nonwords from Pagliuca et al.’s study (2008), 60 bisyllabic nonwords from Arduino et al.’s study (2004), and 32 trisyllabic nonwords taken from Burani and collaborator’s study (2008). The nonwords extracted from the Pagliuca et al. (2008) study were all bisyllabic nonwords, four to six letters long, derived from high-frequency and low-frequency words, by changing at least one letter (or in most cases 2) of the original word. The set of stimuli used by Arduino et al. (2004) contained five to six letters long bisyllabic nonwords. The stimuli used in Burani et al. (2008) were all trisyllabic nonwords, half of which composed of a root (e.g., *donn-*, ‘woman’) plus a derivational suffix (e.g., *–ista*, ‘-ist’) resulting in a combination not existent in Italian (e.g., *donnista*, ‘womanist’), and half were simple nonwords (e.g., *dennosto*) which did not include any existing morpheme. The roots were of high frequency and suffixes were among the most frequent and productive in Italian nominal and adjectival derivatives. The two nonword sets were matched for initial phoneme, syllabic structure, length, bigram frequency, orthographic neighborhood size and orthographic complexity.

The model was successful in reading correctly 98% of the nonwords, which is a level comparable to human performance (Pagliuca et al., 2008). The model therefore shows a good level of generalization to novel stimuli.

### *Frequency Effect*

In addition to the performance on words and nonwords, one of the benchmark effects that every computational model of reading should simulate successfully is the frequency effect. The frequency effect has been documented for all orthographies studied so far, from deep to shallow, including Italian, and has proven to be the most robust finding, with frequency being the psycholinguistic variable that accounts for the largest portion of variance in naming reaction times (Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004). The model was trained 4 times with random initial weights and then each of the four simulations was tested on the 48 words (24 high frequency words and 24 low frequency words) from Pagliuca et al.'s study (2008). These stimuli do not contain any context-dependent rule and each letter in each word entails a perfect one to one mapping with phonology. The stimuli were also matched for the first 2 phonemes, length and summed bigram frequency. A Linear Mixed Effects analysis (Baayen, 2007) was run on the data, with frequency as a fixed effect, number of simulations and stimuli as random effect, and SSE as the dependent variable. The model showed sensitivity to frequency for completely shallow Italian words, with high frequency words having lower SSE than low frequency words ($F(1, 47.49) = 7.83$, $p<.01$, see Figure 2). The model confirms the frequency effect for Italian, an effect that has been taken as evidence for a lexical route in the reading system, even for transparent orthographies. We return to this point in the discussion below, in terms of the extent to which connectionist models can illuminate the debate on whether such a lexical route is necessary to simulate reading behaviour.

*Morphological effect*

The study conducted by Burani and collaborators (2008) sheds light on the sensitivity to large sublexical clusters that Italian readers develop in the course of learning to map orthography onto phonology. A PDP model of reading with no separate mechanisms for reading words and nonwords should still show sensitivity to large grain sizes that are shared by many words in the corpus. That should equally apply to a model that is trained with an extremely shallow orthography such as Italian. Moreover, a model with no semantic pathway should provide support to the hypothesis that morphological effects in naming aloud Italian are non-semantic in nature (Burani et al., 1999).

The two sets of 16 three-syllable pseudowords, morphologically complex (*donnista*) and simple (*dennosto*) from Burani et al. (2008) were selected. Errors accounted for 3.9% of all responses. The model was tested as before, with four runs to simulate 4 "participants". A Linear Mixed Effect analysis was conducted on the SSE as a dependent variable, morphological status as a fixed effect, number of simulations and stimuli as random effects. As Figure 3 shows, the model developed sensitivity to large "morphological" units, and performed better when tested with morphologically complex words than with control words, $F(1, 29.98) = 4.1$, $p = .05$.

*Neighbourhood effect*

The study conducted by Arduino and Burani (2004) provides helpful insights on the ability that readers of a shallow orthography as Italian have to discover and make efficient use of multiple grain sizes, larger than the single letter or bigram. In their study, the authors show that nonwords which share a large cohort of neighbours (or Nsize, as classically describe by McCann & Besner, 1987) are named faster than nonwords with a small cohort of neighbours, irrespective of the frequency of these words (Arduino and Burani, 2004). The effect has been ascribed to the supposed interaction between a lexical mechanism and a nonlexical one in reading these nonwords, with the lexical route boosting reaction times for nonwords that share many lexical neighbours, but not for nonwords with few neighbours. A different view on this effect is adopted by distributed approaches to reading, which assumes the reading system to be fully inherently interactive, with no partitions or separate modules for words and nonwords. A PDP model of reading with no separate mechanisms for reading words and nonwords should still show sensitivity to grain sizes that are shared by many words in the corpus. That should equally apply to a model that is trained with an extremely shallow orthography such as Italian.

The 32 bisyllabic nonwords from Arduino and Burani (2004) were selected. Half of these nonwords (16) have a large Nsize (as the nonword *bento*, which has many lexical neighbours that vary only by the first letter, i.e. *vento*, *sento*, *cento*, *lento*, *pento* etc.), while half (16) a low Nsize (the nonword *biore*, which only has the word *fiore* as neighbour). The model was tested as before, with 4 runs to simulate 4 "participants". A Linear Mixed Effect Analysis was conducted with SSE as the dependent variable, neighbourhood size and frequency as the fixed effect and stimuli and simulations as random effects. The effect of neighbourhood size was not significant ($F(1, 53.5) = 0.13$, ns).

The lack of an Nsize effect could be ascribed to the necessity that the model has to be exposed to the corpus of words for a reasonable amount of time in order to capture subtle effects such as the Nsize effect for a specific dataset of nonwords. Even if the model does not show sensitivity to the nonwords used, it could still show an overall effect of Nsize on the whole corpus. In order to test this hypothesis a multiple regression analysis was carried on the whole corpus of words at the item level, with the Nsize, frequency and bigram frequency used as predictors of SSE value. As table 1 shows, the model is sensitive to word Nsize above and beyond variables such as frequency and bigram frequency, with words with large Nsize having lower SSE than words with a low Nsize. These results suggests that the model is sensitive to the number of neighbours and is processing information coming from N grams larger than 2-3 letters.

One possible explanation of the lack of N size for the model on the behavioural data might have to do with the extent of training needed for the model to fully capture these large orthographic clusters. It is possible that the number of iterations the model was trained with did not suffice and did not enable the model to develop the sensitivity needed to capture this subtle effect. To test this hypothesis, training was extended from 1.2 million to 2 million repetitions, with the same settings described above and the model was retested with the new weights on the same testing material.

This time the model showed a marginally significant effect of neighbourhood effect size for the Arduino and Burani (2004) material, with nonwords belonging to a large cohort of neighbours having lower SSE than nonwords belonging to a small cohort of neighbours ($F(1, 55.7) = 3.64$, $p=.06$, see figure 4).

*Stress assignment*

The study by Burani and Arduino (2004) argues that stress assignment is not generated through the application of a default rule, as suggested by Colombo (1992), but instead is modulated by the number of words which share the same ending and are consistent with the target word's stress pattern. In this light, the word endings used as cues to address stress assignment might be thought of as shared grain sizes between orthographic information and stress patter information. The model we introduced in this paper doe not implement a set of rules to assign stress, and is suitable for testing whether such a rule mechanism is necessary to assign stress to Italian polysyllabic words and whether a connectionist network can show sensitivity to the cues that help adult readers in determining the appropriate stress assignment. Such a simulation would then provide a useful test for the stress consistency versus stress regularity hypothesis.

From the Burani and Arduino (2004) study we selected all the stimuli from experiment 1, with the exclusion of 8 four syllables words, which the model can not process due to architectural constraints. The stimuli were divided in four groups: regular stress with many friends, regular stress with many enemies, irregular stress with many friends, irregular stress with many enemies.

After training the model did not produce any orthographic error for this subset of words, but for 27% of the stimuli it generated a stress patter which did not match the target. These mismatches were considered as stress assignment errors and were excluded from the analysis of sum squared error, and analysed separately. A Linear Mixed Effect Analysis was conducted with SSE as the dependent variable, stress neighbourhood size and regularity as the fixed effect and stimuli and simulations as random effects. The effect of stress neighbourhood size was not significant ($F(1, 34.3) = 0.081$, ns) nor was the effect of regularity ($F(1, 34.3) = 2.094$, ns).

A separate Linear Mixed Effect Analysis was conducted with number of errors as the dependent variable, stress neighbourhood size and regularity as the fixed effect and stimuli and simulations as random effects. The effect of stress neighbourhood size was not significant ($F(1, 44) = 0.256$, ns) nor was the effect of regularity ($F(1, 44) = 0.529$, ns), but there was a significant interaction between regularity and stress neighbourhood size ($F(1, 44) = 4.52$, $p<0.05$), with post hoc analysis revealing that irregular words with many stress friends were read significantly more accurately than irregular words with many stress enemies.

The model of reading that has been presented here inherits the structure, the properties, but also the limitations of classical PDP models of reading (Harm & Seidenberg, 1999, 2004). One of these limitations consists in the syllabic parsing that has been imposed to the model. The slot based representation here employed carries over the so called *dispersion problem* (Plaut et al., 1996). For instance, a phoneme in an initial consonant is represented separately from the same phoneme in a final consonant, and the model has to learn both of them independently. This problem extends to all phonemes within a syllable, and in our case it extends beyond it, encompassing three syllabic sets of slots, making the task of learning phonological representations hard for the model.

However, as noted by Harm and Seidenberg (1999), the presence of the connections between the phonological layer and the set of clean-up units enables the model to capture dependencies across different slots, and it also enables to model to capture the fact that phonemes in different positions behave differently. Interestingly, the question remains as to whether a model with no syllabic parsing would still be able to detect large grain sizes in a transparent orthography as Italian. It is also unclear whether centering the phonological code around the vowel had any impact on performance in our model. For English, vowels represent the major source of phonological variability in the orthographic-phonological mapping and centering around the vowel greatly helps in reducing the inconsistencies in the mapping. For Italian, this case is largely reduce, due to the strong consistency of the mapping itself. However, the introduction of stressed vowels in our phonological representation might have increased the level of inconsistency and the centering might have helped our model as well. There is therefore a possibility that the syllabic parsing we employed helped the model to detect the set of effects we simulated, and that a model with no such syllabic encoding would have failed to do so and would have limited its sensitivity to small grain sizes only (letters or bigrams), only in specific positions.

Following we discuss a model of reading for Italian that does not employ syllabic parsing in the orthographic representation for the second syllable, yet it is capable of showing sensitivity to grain sizes that span more than one syllable.

## Simulation 2: Modelling Italian without orthographic parsing

*Method*

*Architecture and Representation*

The architecture is largely based on the previous model, with one differences in the way orthography was encoded. At the orthographic layer, instead of centering each syllable around the vowel, we centered only the first vowel of the word on the fourth slot. The second syllable was then left aligned to the first, with no parsing or empty slots separating the two syllables. A word as *babbo* (dad) was then represented as - - b a b b o - -, and a word as *fabbro* (smith) was inputted as - - f a b b r o -. In this case, the model has no way to know that the final letter "o" is the same in both words, as it occupies a different slot each time. Similarly, the model does not know where the second syllable starts, as there are no separators or empty slots between the first and second syllables. This type of representation greatly increasing the difficulty of learning the mapping between orthographic input and phonological output. Given the difficulty of the task, we only trained and tested the model with mono and bisyllabic words. The orthographic layer comprised 252 units, the hidden layer had 100 units and the phonological layer contained 88 units. A set of 50 cleanup units was added to the network and connected bidirectionally to the phonological layer in order to create a set of phonological attractor units, as before. The phonological representation used was the same as for the original model, whereby parsing was syllabic based at the output level. Instead of having the model generate the entire phonological output at the same time slice as for the previous version, we had it produce one syllable at a time (the first and then the second) using the same phonological units. At time ticks 10-11 the model was generating the first syllable, and at time ticks 22-23 the second syllable. This way of generating the output partially gets around the classic parallel distributed problem of producing the first and the last phonemes in a word at the same time, which for a polysyllabic word is implausible. However, it is important to note that the orthographic information was still inputted in parallel, with all the orthographic slots for a given word being activate at the same time.

*Training corpus and regime*

A total of 2424 mono and bi-syllabic words were used as training corpus. Frequency was compressed as before. Given the difficulty of the task the model was trained for 10 million iterations. In order to save on computational resources, the phonological attractor was not pretrained, but was used online during training and testing. The training parameters were the same as before.

*Results and Discussion*

Naming accuracy and sum squared error (SSE) were computed to test the model's general performance, as before. After 10 million word presentations the model could read correctly 96% of all words. Orthographic errors accounted for 2% and the remaining 2% were classified as stress placement errors (for bi-syllabic words stressed on the second syllable, the model failed to stress the last vowel).

The main question that led to the development of this model was whether a model with no orthographic syllabic parsing could still show sensitivity to large grain sizes that span more than one syllable. We tested the model on the neighbourhood effect (Arduino et al., 2004). We chose this test because the current model was only trained on bi-syllabic words, and the test stimuli are all bi-syllabic. A main effect of neighbourhood size was again found ($F(1, 56.2) = 7.09$, $p<.05$) with nonwords belonging to a large cohort of neighbours having lower SSE than nonwords belonging to a small cohort of neighbours. The model is showing sensitivity to large grain sizes (word neighbours) that encompass more than one syllable, even with no explicit syllabic parsing employed at the input layer.

The type of encoding employed in this simulation however is not immune to limitations. The model was tested on a subset of 108 bi-syllabic nonwords taken from the corpus used in the first version of the model. The model could read correctly 87.4% of these nonwords, which falls well below human performance on the same stimuli (>98%). This poor level of performance highlights the need to adopt an encoding scheme that facilitates the process of identifying syllabic boundaries and syllabic structures. However, the lack of such an encoding scheme does not seem to have an effect on the model's ability to detect different grain sizes for a transparent orthography as Italian.

## General Discussion

This paper introduced a connectionist PDP model of reading for Italian. The model inherits all the properties of standard distributed models of reading and extends the reach of this class of architectures to a transparent orthography and to a large corpus of polysyllabic words. One of the main advantages of our model is the ability to effectively represent polysyllabic words and their stress pattern with a standard PDP model of reading. The model learned to read 98% of the Italian lexicon accurately, even though words varied in length from one to three syllables. The model had to learn, then, when the orthographic input referred to a single syllable, and when it should be divided up to contribute to producing multiple syllables. The model, therefore, showed that polysyllabic reading models are within the remit of such connectionist models, without recourse to the complex orthographic recoding as used by the Ans et al. (1998) model to achieve polysyllabic reading. The model was also successful in generalising its reading behaviour to nonwords, indicating that it had learned, not just the specific mappings of the lexicon, but general statistical relationships between letters and phonemes to enable a broad range of stimuli to be named.

Despite the extreme regularity of the mapping between Italian orthography and phonology, the model managed to show sensitivity to grain sizes larger than the single unigram or bigram and to capture subtle effects involving the use of large orthographic and phonological clusters, effects that have been documented in several behavioral studies (Burani et al., 2008; Pagliuca et al., 2008), at the lexical and sublexical (morphological) level. The model, in line with classic PDP architectures, does so with no explicit localist representation of these large grain sizes (lexical and/or morphological units) and with no recourse to an explicit lookup mechanism to the lexicon, as employed in dual route models of reading (Perry, Ziegler, & Zorzi, 2007). More importantly the model does not explicitly implement a set of grapheme to phoneme conversion rules (Coltheart et al., 2001), but it learns the relationships in the mapping during training, relationships that go beyond the single letter-phoneme mapping to encompass a wide range of grain sizes.

Despite the strict and consistent one-to-one correspondence between the input (letters) and the output (phonemes), the model does not learn to rely exclusively on small grain sizes, a coding strategy that alone would prove successful in reading most words in Italian, but learns a whole range of grain sizes, in some cases as large as the morpheme, and takes advantage of these shared clusters in generating the output, as both Italian experienced and young readers do as shown in the behavioural studies on reading in Italian (Arduino et al., 2004; Burani et al., 2008; Pagliuca et al., 2008). The model also suggests along with the behavioural studies (Burani et al., 2008) that morphemes are useful shared grain sizes that should be considered as functional reading units at least in morphologically complex orthographies such as Italian.

The model was effective in replicating the morphological effects without recourse to a semantic system, or an explicit lexical or sublexical system that encodes at the level of the morpheme. Rather the model discovers the relevant unit in terms of mapping between orthography and phonology. As morphemes occur frequently in the lexicon, this has the effect that certain patterns of letters co-occur frequently, and map reliably onto a set of phonemes. The connectionist model is sensitive to such regularities, and can encode all levels of granularity that are useful for forming the mapping.

Similarly the model showed sensitivity to orthographic neighbourhood size (Arduino & Burani, 2004), with nonwords with many neighbours being advantaged over nonwords with few neighbours. As with the morpheme, the shared grain size represented by the orthographic neighbours was successfully detected and provided and advantage in processing words which belonged to a large cohort of neighbours. This effect was replicated with a version of the model which did not implement orthographic parsing, thus showing that the imposed input structure had little effect on the model's ability to detect different grain sizes.

The model was also tested in order to explore sensitivity to another grain size, represented by the word ending used as a cue for stress assignment in Italian (Burani & Arduino, 2004). No such effect of stress consistency was detected, nor a main effect of regularity, as in the Colombo's study (1992). However, the model showed the same interaction found by Colombo (1992) between regularity and number of stress neighbours, with irregular words with many friends being read more accurately than irregular words with many enemies. Due to the lack of a stress assignment default mechanism in the model, these results do support the claim that such a rule system is necessary to correctly assign stress in Italian. The lack of a consistency effect as in the Burani and Arduino (2004) study could highlight the limitations of the current architecture of the model, and perhaps the corpus used to train it. In the model stress has been represented as feature within the phonological space, and not a suprasegmental feature as classically thought in general linguistics. The absence of a large corpus of four syllable words might have also hampered performance by reducing the model's exposure to a larger set of stress friends and enemies in richer orthographic clusters within longer words. These limitations might prompt more work in the direction of building a model that can deal effectively with long polysyllabic words.

In short, the model here presented shows that parallel distributed approaches to reading can prove a powerful tool to explore the reading system not just for deep orthographies like English, but for transparent orthographies as well. We have argued that the grain size for a language may be variable, and is a property of generalised features of the lexicon as well as the particular configuration of the word in question. Though Italian is more transparent in the orthography-phonology mapping than English, still multiple letter units, and variably grain sizes, are observed and explained by a connectionist model that reflects the statistical properties of a representative lexicon of the Italian language.

## Acknowledgements

This work was supported by an EU Sixth Framework Marie Curie Research Training Network Program in Language and Brain: http://www.ynic.york.ac.uk/rtn-lab. The second author was supported by ESRC grant RES-000-22-1975. We are indebted to Jangfeng Yang for invaluable support provided during the development of the model, to Cristina Burani and two anonymous reviewers for helpful comments on an earlier version of the draft.

An executable version of the model can be downloaded from:

http://www.york.ac.uk/depts/psych/www/people/biogs/gp517.html .

*References*


Ans, B., Carbonnel, S., Valdois, S. (1998). A connectionist multiple-trace memory model for polysyllabic word reading. *Psychological Review*, *105*, 678-723.

Arduino, L.S., & Burani, C. (2004). Neighborhood effects on nonword visual processing in a language with shallow orthography. *Journal of Psycholinguistic Research, 33,* 75-95.

Baayen, R.H. (2007). *Analyzing linguistic data: A practical introduction to statistics.* Cambridge: Cambridge University Press (in press).

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

Bertinetto, P.M., Burani, C., Laudanna, A., Marconi, L., Ratti, D., Rolando, C., & Thornton, A.M (2005). *Corpus e Lessico di frequenza dell'italiano scritto (CoLFIS)* [lexical database].

Burani, C., Dovetto, F.M., Spuntarelli, A., & Thornton, A.M. (1999). Morpholexical access in naming: the semantic interpretability of new root-suffix combinations. *Brain and Language, 68,* 333-339.

Burani, C., & Arduino, L.S.(2004). Stress regularity or consistency? Reading aloud Italian polysyllables with different stress patterns. *Brain and Language*, *90*, 318-325.

Burani, C., Barca, L., & Ellis, A.W. (2006). Orthographic complexity and word naming in Italian: some words are more transparent than others. *Psychonomic Bulletin and Review, 13,* 346-352.

Burani, C., Marcolini, S., De Luca, & Zoccolotti, P. (2008). Morpheme-based reading aloud: evidence from dyslexic and skilled Italian readers. *Cognition, 108,* 243-262.

Canepari, L. (1980). *Italiano standard e pronounce regionali*, Padova, Cleup.

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.

De Mauro, T. (2000). Il Dizionario della Lingua Italiana, Paravia.

Harm, M.W., & Seidenberg, M.S. (1999). Phonology, reading and dyslexia: insights from connectionist models. *Psychological Review*, *163*, 491-528.

Frost, R., Katz, L., & Bentin, S. (1987). Strategies for visual word recognition and orthographical depth: a multilingual comparison. *Journal of Experimental Psychology: Human Perception and Performance*, 13, 104-115.

Harm, M.W., & Seidenberg, M.S. (2004). Computing the meaning of words in reading: cooperative division of labor between visual and phonological processes. *Psychological Review*, *111*, 662-720.

Hutzler, F., Ziegler, J. C., Perry, C., Wimmer, H., & Zorzi, M. (2004). Do current connectionist learning models account for reading development in different languages? *Cognition, 91*, 273-296.

McCann, R., & 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.

Monaghan, P. & Christiansen, M.H. (2006). Why form-meaning mappings are not entirely arbitrary in language. *Proceedings of the 28th Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Lawrence Erlbaum Associates.

Pagliuca, G., & Monaghan, P. (2008). Variable vulnerability of words to visual impairment: exploring grain-size effects in reading. *Proceedings of the Thirtieth Annual Conference of the Cognitive Science Society.* Mahwah, NJ: Lawrence Erlbaum (2008).

Pagliuca, G., Arduino, L.S., Barca, L., & Burani, C. (2008). Fully transparent orthography, yet lexical reading aloud: The lexicality effect in Italian. *Language and Cognitive Processes*, 23, 422-433.

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.

Plaut, D.C., & Gonnerman, L. M. (2000). Are non-semantic morphological effects incompatible with a distributed connectionist approach to lexical processing? *Language and Cognitive Processes*, *15*, 445-485.

de Saussure, F. (1916). *Cours de linguistique générale*. Paris: Payot

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

Seymour, P.H.K., Aro, M., & Erskine, J.M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology*, *94*, 143-174.

Thornton, A.M., Iacobini, C., & Burani, C. (1997). *Una base di dati per il vocabolario di base della lingua italiana* (seconda edizione riveduta e ampliata). Roma: Bulzoni.

Zevin, J.D., & Seidenberg, M.S. (2006). Simulating consistency effect and individual differences in nonword naming: a comparison of current models. *Journal of Memory and Language*, *54*, 145-160.

Ziegler, J., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin, 131,* 3–29.

Table 1. Hierarchical Regression on SSE.

| Variable | Adjusted $R^2$ | B | t | p |
|---|---|---|---|---|
| Step 1 | | | | |
| Frequency | 0.14 | -.117 | -11.77 | .000 |
| Step 2 | | | | |
| Bigram Frequency | 0.17 | -.061 | -6.17 | .000 |
| Step 3 | | | | |
| Nsize | 0.19 | -.047 | -4.65 | .000 |

List of Figures.

Figure 1.

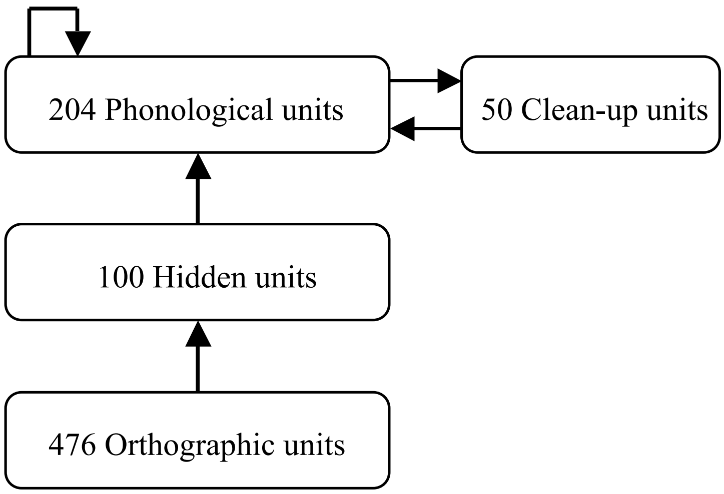
204 Phonological units
50 Clean-up units
100 Hidden units
476 Orthographic units

Figure 2.

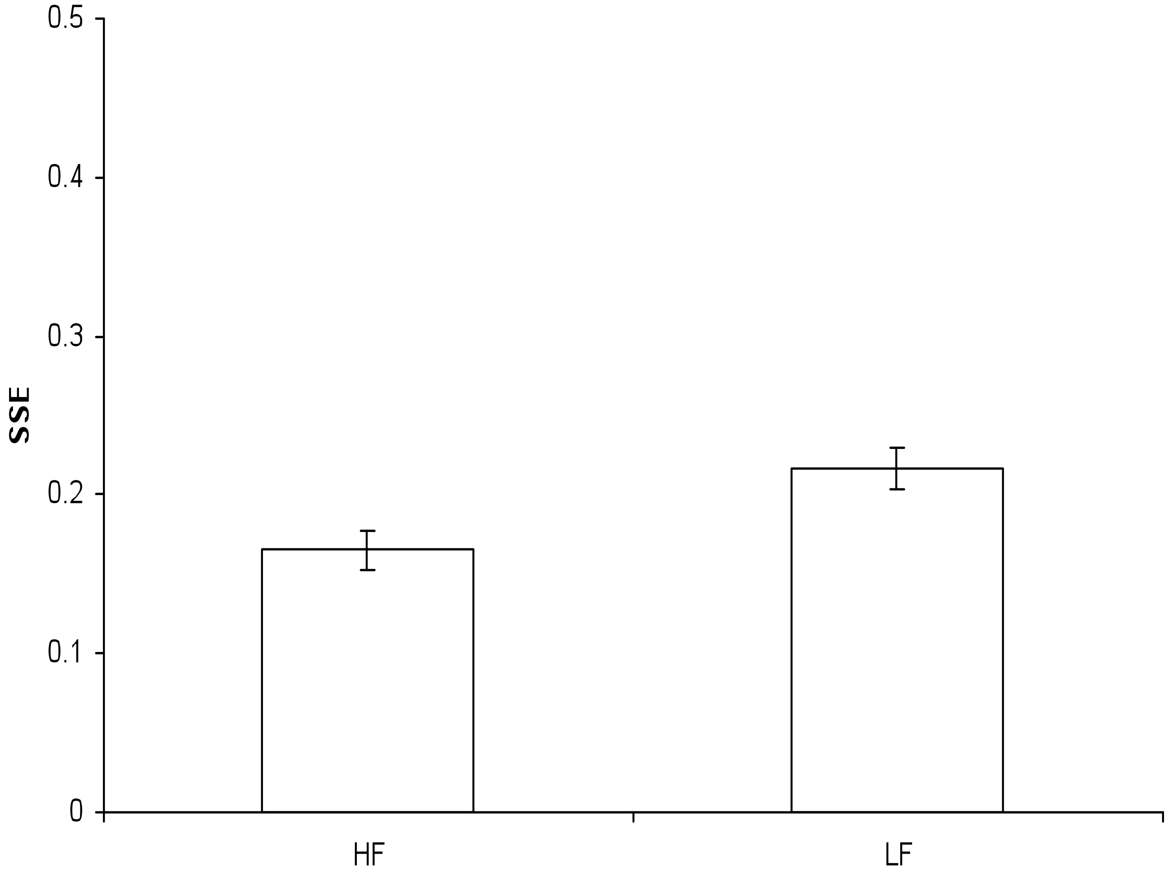

Figure 3.

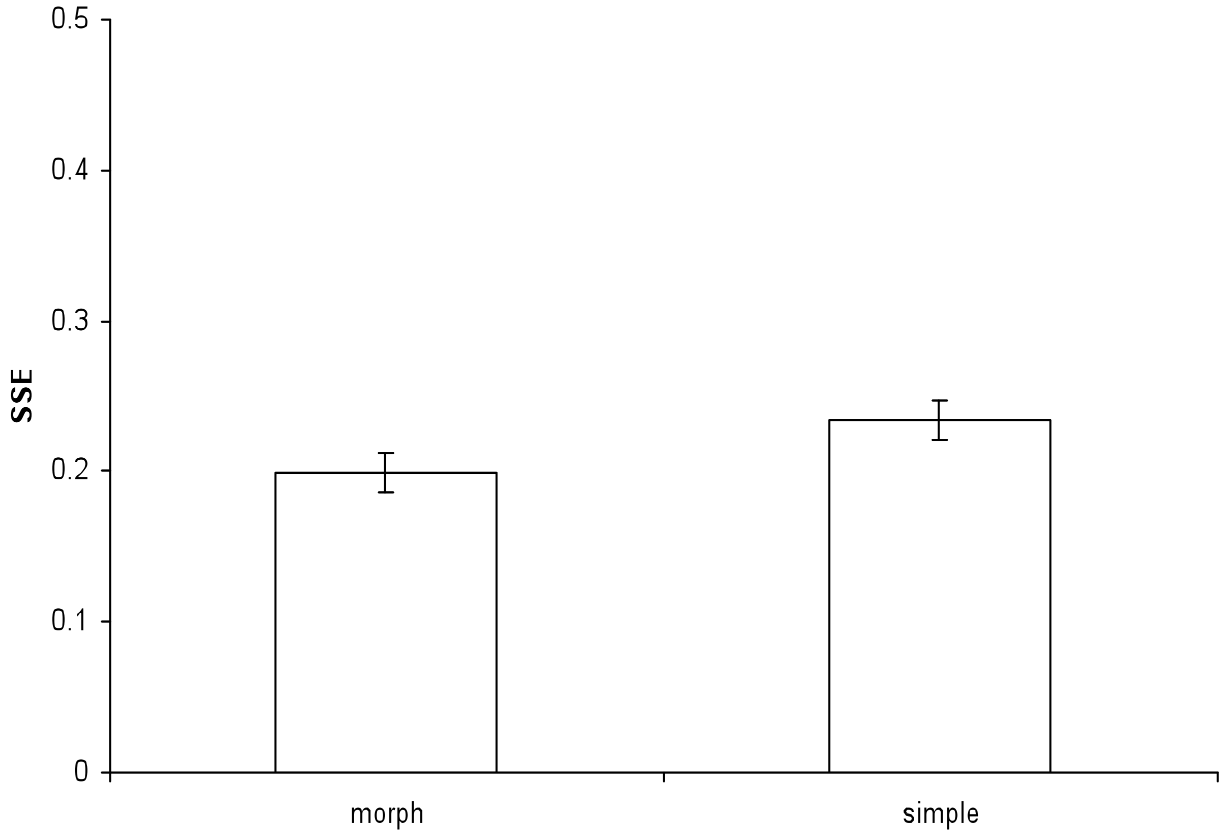

Figure 4.