universität wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

"Investigating lexico-semantic access of pseudohomophones with a picture task: implications for models of reading aloud"

verfasst von / submitted by

Andrea Hackl, BA

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Arts (MA)

Wien, 2018 / Vienna, 2018

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	A 066 812
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium English Language and Linguistics UG2002
Betreut von / Supervisor:	Univ.-Prof. Mag. Dr. Nikolaus Ritt

Acknowledgments

Having arrived at the end of this research project, I would like to take the time and thank those who have supported me along the way. First, I would like to thank Professor Ritt for accepting such an unusual topic in the first place – you really approach new ideas with an open mind and support young researchers in their efforts. I would also like to thank Dr. Andreas Gartus, who works at the Department of Basic Psychological Research and Research Methods, for his valuable suggestions on how to conduct a cognitive experiment like the one I had in mind with regard to equipment and software. Thanks goes out to Professor Gary Libben for reviewing my research proposal and offering useful comments, and to Professor Mario Braun for suggesting literature on the necessity to control for initial phoneme in psychological experiments. And since most experiments would not be possible without participants, I sincerely thank everyone who took the time to be a part of this research project.

On a more personal note, I would like to thank my husband Mike for his love and encouragement, for believing in me and for pointing out new ways. I would also like to thank Ina Rager, who is the most diligent study partner that anyone could wish for, and has come to be a good friend. Thanks to Kathrin Wiesinger for agreeing to do a pub survey on Viennese monophthongization for our linguistics proseminar – and many other pub ‘surveys’. Welcome distractions were also provided by my pub quiz team members Mario I, Mario II, Martina, Tina, and Helmut, who are a smart and lovable bunch of people. And finally, though it may seem a bit unconventional, I would like to thank myself for being so tough and hard-working, and for always pushing through.

Table of contents

List of abbreviations

CDP+	Connectionist dual-process plus model
DRC	Dual-route cascaded model of visual word recognition and reading aloud
GPC	Grapheme-phoneme conversion
PDP	Parallel distributed processing model of visual word recognition and pronunciation
PH(s)	Pseudohomophone(s)
TLA	Two layer associative network

List of figures

List of tables

1. Introduction

As pointed out by Seidenberg (2007: 241), "reading [is] a cultural artefact created very recently in human history, [and] makes use of capacities (language, vision, learning, thinking) that evolved for other purposes". In fact, reading is a skill that we use on a daily basis to access information that is made available to us in writing. In their early years, children acquire language based on listening to speech, and once these phonological and grammatical patterns have become established, they are confronted with the need to decipher written codes in order to read their school textbooks. The process by which children learn to read letters is often referred to as 'sounding out': children use spelling-sound relationships to translate graphemes[1] into their corresponding phonemes. This way, familiar phonological patterns enable the child to identify a word and, ideally, to understand the meaning behind the unknown letter combination.

By the time they reach adulthood, grown-ups will have encountered thousands of words that they are able to read without difficulty. But adult native speakers also exhibit remarkable skill at reading aloud words which they have never seen before, or even words that do not exist. How this process is accomplished is the subject of a longstanding debate, whereby two main approaches to reading have been taken. One approach stipulates that humans possess two distinct routes for reading aloud known and unknown words (dual-route models, e.g., Coltheart et al. 2001), the second approach argues that there is a single process that can account for reading aloud both types of words (most notably the PDP approach by Seidenberg & McClelland 1989). This interest in the cognitive structures underlying word reading resulted in the creation of models of reading aloud, at first with the help of box-and-arrow diagrams, which after the advent of information technology could be made more explicit through the programming of computational models.

In models of reading aloud, researchers are primarily interested in the conversion of letters to sound. The process of meaning-making is typically a byproduct of reading aloud and thus not a major concern. In general, it is assumed that the role of semantics in reading aloud is rather unimportant, and few efforts have been made to develop an integrative approach that includes orthography, phonology, and semantics. Seeing that the influence of semantics on reading aloud has received little attention in the literature, this paper will investigate the role of semantics in

[1] Coltheart (2005: 198) defines 'grapheme' as "a letter sequence that represents a single phoneme". The word *night* has 3 graphemes: 'n', 'igh', and 't', as the letter sequence 'igh' is a grapheme that represents a single phoneme.

three models of reading aloud; moreover, it will look at different possibilities of conceptualizing semantics for this purpose. It should be pointed out that a possible contribution of semantics to reading aloud is least relevant for unknown words or nonwords. Still, unknown words could trigger associations of similarly spelled words, and there is one special case in which semantic access could be hypothesized to take place, namely in the reading of pseudohomophones (PHs). PHs (e.g., *brane* or *nale)* are words whose orthographic representation is unknown to the reader, but whose phonological representation corresponds to an existing lexical entry (i.e., *brain*, *nail*).

The present paper focuses on this special case of PHs. Experiments conducted in this area usually include speeded naming or lexical decision tasks (cf., Section 4 for an overview of commonly used methods). In lexical decision, participants are asked if a word is a real word; in the case of a PH, the correct answer is 'no'. Still, a PH shares the phonology of a real word, and is endowed with meaning through its sound pattern. But meaning is typically neglected in these experiments – researchers devise all sorts of methods, but hardly any of these include semantics. In this thesis, I aim to fill this gap by taking a closer look at the role of semantics in PH reading. To be more precise, my overall research question is the following: To what extent is meaning activated when reading aloud PHs? This question is investigated through the combination of a naming and a picture task. The first task requires participants to read aloud a mixed list of words and PHs; after completing this task, participants are asked which words they remember. The second task is a forced-choice picture task in which some pictures resemble objects that have been read aloud, whereas others do not. Pictures are either a match, semantically related (e.g., *egg* / chicken) or semantically unrelated (e.g., *beer* / fan). In addition to correctness of response, response times are collected.

Through foregrounding semantics, this paper represents a linguistic take on the topic of reading aloud unfamiliar words. After outlining the basic architecture of three models of reading aloud as well as the role of semantics in these models, the PH effect as well as findings in this area will be presented. Strictly speaking, this literature review might be more extensive than it would need to be in order for a reader to understand the research project. However, I believe that this topic area lacks a concise overview in the existing body of literature, and I therefore attempt to generate such an overview that future researchers can turn to as a basis for their own research. In Section 5, I discuss possible approaches to conceptualizing semantics for the purpose at hand; these approaches are Saussure's (1986) 'signifier'/'signified', Frege's (1948) 'sense' and 'reference', and Miller's (1995) semantic database WordNet. My research suggests that semantic access in PH reading takes place to a large extent, at least when highly imageable PHs (that can be easily pictured in the mind) are mixed with real words. Participants remember more PHs than real words, and in the picture task, errors are few in general. Implications for models of reading aloud will be discussed.

2. On words and nonwords

According to the OED (accessed 10 Jul 2017), a non-word is defined as “an unrecorded or hitherto unused word, a word which has […] no accepted meaning”. In fact, nonwords have been used for a long time and for different purposes. In *Through the looking glass* (Carroll 1872), the second part of *Alice’s adventures in wonderland* (Carroll 1865), Lewis Carroll included a poem about the killing of a fictional creature called the *jabberwocky*; this poem consists of nonwords to a large extent: “Beware the Jabberwock, my son! / The jaws that bite, the claws that catch! / Beware the Jubjub bird, and shun / The frumious Bandersnatch!” (Carroll 1998[1872]: 132). Clearly, this use of nonwords serves the purpose of artistic expression. An example of a linguistic reason for the use of nonwords is the ‘wug test’ created in 1958 by Gleason. This test was developed to see what grammatical rules a child had already acquired. The child is shown a picture of an unknown creature and learns that it is called a *wug*. Then the child is asked what it would call two of these creatures – a simple test of plural formation. Another linguistic purpose is exemplified by the V_YesNo test (Meara & Miralpeix 2015). This is a test of receptive vocabulary in which the test taker is asked if s/he knows the meaning of single words presented in isolation. If the participant knows the word, s/he clicks ‘yes’, if s/he does not know the word, s/he clicks ‘no’. To ensure that the participant does not simply click ‘yes’ in response to every word, words are mixed with nonwords, and if the ‘yes’ option is selected in case of a nonword, points are deducted.

In psychological research, nonwords are used with the aim of deepening our understanding of how language is processed by the human brain. This usage of nonwords is particularly relevant to the area of reading (aloud). In researching the process of reading aloud, three types of words have typically been distinguished: real words, nonwords and PHs. The notion of ‘real word’ is ususally taken for granted; the OED (accessed 10 Jul 2017) defines ‘word’ as “[s]omething that is or has been said; an utterance, a statement, a speech, a remark”. However, this definition is imprecise. In order to be considered a real, existing word, a letter combination should ideally be known by the vast majority of native speakers, and in the case of a standardized language, it should have been recorded in a dictionary. If these criteria do not apply, it is probably more appropriate to consider the letter combination a nonword. Such nonwords can either be pronounceable or unpronounceable; pronounceable nonwords adhere to the phonotactic restrictions of a language and are generally referred to as pseudowords.

Overall, a language may possess a transparent (shallow) or an opaque (deep) orthography. This means that its grapheme-phoneme correspondences are largely systematic (e.g., Spanish) or irregular (e.g., French), in other words, the pronunciation of a word can be readily derived from the letters that are used to spell it. In an opaque orthography such as English, real words are generally divided into regular and irregular/exception words, depending on the regularity of their spelling-sound correspondences (cf., Glushko 1979). If the pronunciation of a word does not conform to the expected 'majority pattern', it is considered irregular (e.g., *have* is irregular as opposed to *gave*). A further distinction is made between consistent and inconsistent words. If a word's body is not always pronounced in the same way in other words containing the same body, the body is said to be inconsistent (Coltheart 2012: 6-7). In this sense, words are said to have friends, i.e., words that share the same body pronunciation (e.g., *wave* and *gave*) and enemies, i.e., words that have a different pronunciation for the same letter combination (e.g., *wave* and *have*). Consistency can either be seen in absolute terms or in terms of gradience in accordance with a word's friend-enemy ratio (cf., Jared 2002).

Another aspect related to pronunciation is homophony. If two words share the same pronunciation, but differ in spelling as well as meaning, they are referred to as homophones (e.g., *flower* and *flour*). In the area of nonwords, homophones have a counterpart called PHs. Whereas the orthographic representation of PHs is unknown to the reader, their phonological representation corresponds to an existing lexical pattern. To take a specific example, the PH *brane* shares its phonology and lexical meaning with its real-word equivalent *brain*. These PHs cannot be recognized based on orthography alone, but skilled readers can assemble their phonology and possibly access the related semantic content. This unique characteristic makes PHs an interesting area of study. Section 4 outlines the methods that have typically been used to study the processing of PHs and summarizes the outcomes of studies in this field.

Figure 1 visualizes the various types of words that have been distinguished in this section.

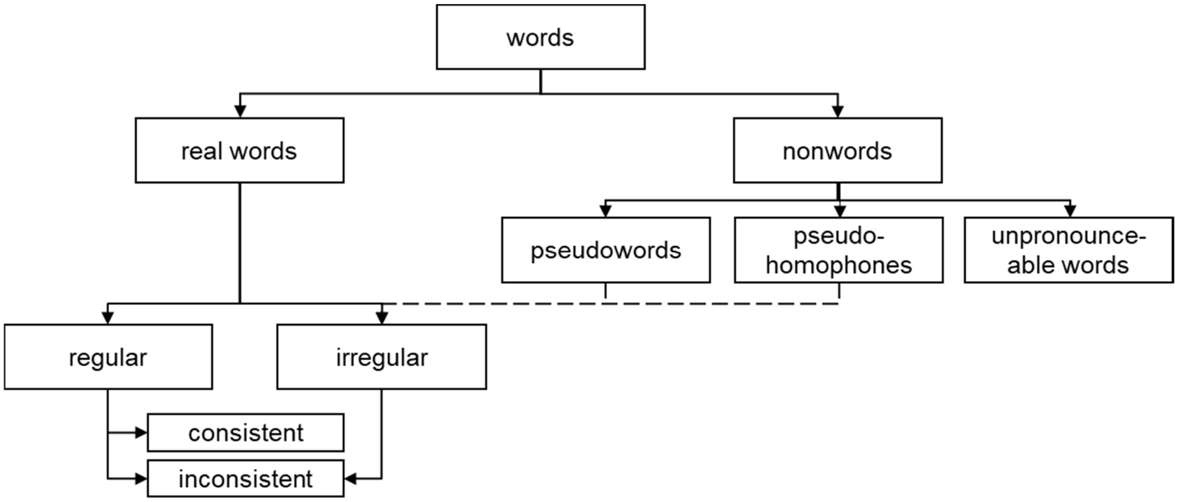


Figure 1 The nature of words and nonwords

On the whole, words can be divided into real words and nonwords; nonwords may further be divided into pseudowords, PHs and unpronounceable words. Another level consists in the notion of regularity (regular versus irregular words) and consistency (consistent versus inconsistent words). The knowledge of these concepts will be necessary for the proper understanding of the following sections. Seeing that the cover term nonword has come to be used instead of pseudoword in most studies, this convention will be adopted in this paper.

3. Computational models of reading aloud

It has long been argued that pseudowords tend to be processed differently from real words. Specifically, Baron (1977: 183) suggested two separate mechanisms for lexical access which he termed "word-specific" and "orthographic". Corresponding models have been referred to as dual-route models of reading aloud (e.g., Forster & Chambers 1973; Coltheart 1978; Patterson & Morton 1985). The underlying idea of these models is that real words are stored in a mental lexicon and can be accessed directly, whereas unfamiliar words and nonwords need to be assembled on the basis of specific rules that translate orthographic signs into their corresponding sounds. This would mean that real words can be taken in as a whole (given enough practice), whereas fictional words are put together one sound at a time. However, this view has been challenged by a number of researchers. To take a specific example, Glushko (1979: 686) argues against a separate mechanism which "employ[s] abstract spelling-to-sound rules". Instead, Glushko (1979: 677) suggests that the pronunciation of pseudowords is largely realized through a phonological comparison with real words, e.g., *hean*/*dean*. Seeing that research participants frequently report the use of analogy as a strategy for pronouncing nonwords (e.g., Glushko 1979; Beech 2010; Woore 2010), it is indeed likely that orthographically similar words or structures are activated along with a pseudoword.

Models of reading aloud have first been developed in the second half of the 20th century. To explain how the task of reading is accomplished, these models were first illustrated with the help of box-and-arrow diagrams. Since, the rise of Information Technology has allowed researchers to develop computational models out of these graphic representations. According to Coltheart (2005: 11), "[a] computational model of some form of cognitive processing is a computer program which not only executes that particular form of processing, but does so in a way that the modeler believes to be also the way in which human beings perform the cognitive task in question". As Busemeyer & Diederich (2010: 6) point out, it is not possible for a computational model to fully mirror the actual cognitive processing performed by humans, it can only be intended as an approximation. The question is, then, which model better matches the known findings from human data.

The present section will introduce three models of reading aloud: the dual-route cascaded model of visual word recognition and reading aloud (DRC) (Coltheart, Rastle, Perry, Langdon & Ziegler 2001), the parallel distributed processing model of visual word recognition and pronunciation (PDP) (Seidenberg & McClelland 1989; Harm & Seidenberg 1999, 2004; Plaut, McClelland, Seidenberg & Patterson 1996), and the connectionist dual-process plus model (CDP+)[2] (Perry, Ziegler & Zorzi 2007). The reason for introducing these three models is that they have been widely discussed in the literature, and all of them have been implemented as computational models, which means that they are fully formulated. To provide a comprehensive account of these models would by far exceed the scope of this paper, and would likely require a much higher expertise than I pretend to possess at this moment. Thus, I will limit myself to presenting the basic architecture of each of the three models and will then move on to discuss the role of semantics in their design. Any claims made with regard to the PH effect and a base-word frequency effect in relation to PHs will be discussed in Section 4.

3.1 Dual-route cascaded model of visual word recognition and reading aloud (DRC)

DRC is based on the first computational model of reading, the Interactive Activation and Competition (IAC) model by McClelland and Rumelhart (1981) and Rumelhart and McClelland (1982), which received widespread attention and approval. Initially, DRC was set up as a box-and-arrow diagram (Figure 2) based on a dual-route architecture. The model was

[2] For more than one syllable, the CDP++ model has been developed.

then turned into a computational model, c.f., Coltheart, Curtis, Atkins & Haller (1993) for an initial approach, Coltheart, Rastle, Perry, Langdon & Ziegler (2001) for an in-depth account, and Coltheart (2005) for an accessible overview.

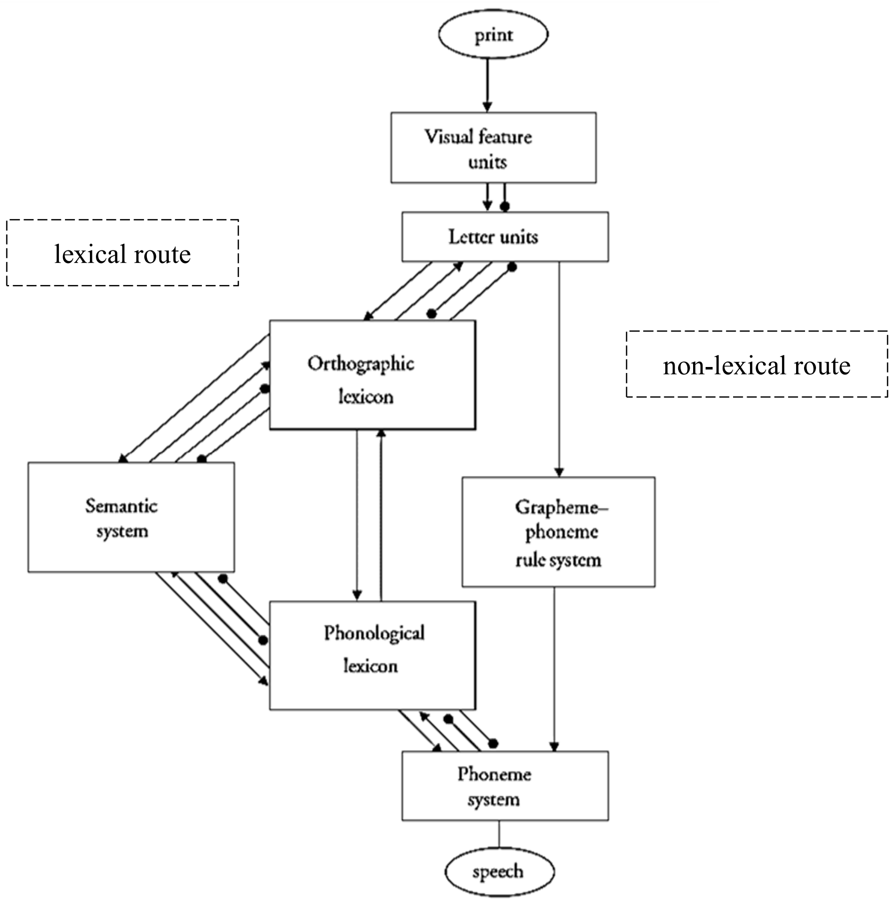


Figure 2 The basic architecture of DRC (Coltheart et al. 2001: 214)
(boxes with dotted lines added)

As Figure 2 illustrates, there is a basic division into the lexical route (a kind of dictionary look-up procedure for familiar words) and the non-lexical route (used for assembly letter-by-letter). Whereas known words are accessed as a whole (their pronunciation is 'retrieved' from an existing lexical entry), rare or unfamiliar words have to be assembled on the basis of conversion rules. The diagram starts with 'print' as input; the letters are analyzed from left to right. Following this analysis, the lexical route starts to look for the letter combination in the orthographic lexicon, whereas the non-lexical route starts to assemble the word one letter at a time. In this respect, it should be mentioned that both routes may contribute to the production of speech. Essentially, they operate in parallel, with the non-lexical route starting to operate slightly later than the lexical route (Coltheart 2005: 14). The fact that both routes may contribute to the end result is frequently explained with reference to the bucket metaphor that was first introduced by Baron (1977: 203): the routes are like hoses that both fill up a bucket, even if they do not contribute an equal amount of water. This involvement of two routes may sometimes lead to differing pronunciations, producing a conflict at the phoneme level. To illustrate this, DRC's non-lexical route yields only regular pronunciations for irregular/exception words, which would result in an incorrect pronunciation unless the lexical route 'wins'. Coltheart (2005: 15) states that such a conflict is resolved "via the interplay of inhibition and activation at various levels of the model", the parameters being set so as to maintain a balance between the routes.

It should be mentioned that DRC is based on the existence of a mental lexicon in which word forms are stored separately from word meanings (Coltheart et al. 2001: 208-209). There is a feedforward from the units in the orthographic to the units in the phonological lexicon, and a feedback from the units in the orthographic lexicon to the letter units. Likewise, there is a feedforward from the units in the phonological lexicon to the phoneme units, and a feedback from the units in the phonological lexicon to the units in the orthographic lexicon (Coltheart 2005: 13-14). If the word is not encountered in the orthographic lexicon, its pronunciation has to be assembled on the basis of grapheme-phoneme conversion (GPC) rules. Overall, DRC uses a cascaded design and localist representations of words; 'cascaded' means that information is passed on from one level to the next. In a cascaded design, there is no threshold level – any amount of activation is sufficient, and processing at one level does not need to be finished before activation spreads to the next level (Cembrani 2010: 5).

The non-lexical route of DRC is based on conversion rules that are perfectly regular in nature. The assembly of unknown words is carried out with the help of a "rule-based algorithm" (Pritchard et al. 2012: 1269); essentially, DRC was programmed with the most common rule for a grapheme, based on type rather than token frequency, and this rule is applied to nonword reading to 100 % (1269). As a result, the reading of nonwords is perfectly regular, as the model never deviates from its established rules. These rules have been programmed (entered into the system), so there is no training database or learning algorithm being applied from which the model could deduce the GPC rules of the language. Due to this characteristic, the route has also been called "hard-wired" (e.g., Pritchard et al. 2012: 1269). In total, there are 236 explicit rules; this number includes 27 context-sensitive rules (e.g., 'c' is interpreted as /s/ before the letters 'e', 'i' or 'y', and as /k/ otherwise) and 8 output rules (largely phonotactic rules, e.g., 'n' is interpreted as /ŋ/ before /k/) (Coltheart et al. 2001: 216; Pritchard et al. 2012: 1269). To explain how these rules work with a specific example, the grapheme /oo/ translates into /u:/, as in *tool*. Theoretically, the grapheme could represent the phoneme /ʊ/ in some words (as in *book*), but DRC only knows the most common rule, and this rule is always applied in pseudoword reading (Pritchard et al. 2012: 1269).

A primary concern of the proponents of DRC was to explain findings obtained from human data. It seems that DRC can explain an unparalleled number of effects found in experiments, e.g., frequency, lexicality[3], and regularity[4] effects (Coltheart et al. 2001: 251). In this respect, Seidenberg & Plaut (2006: 30) argue that Coltheart et al. (2001) focus on the findings of select studies only, while ignoring those of others. Pritchard et al. (2012: 1272) stress the superior performance of DRC in the output of nonword pronunciations when using the responses of human participants as a benchmark. However, the assessment depends on the criteria that are used to evaluate the model's performance. If the model only needs to produce a pronunciation that was given by at least one participant, DRC is sure to perform well, since some of the participants are bound to give the most regular pronunciation in reading aloud a word, and the nonword output of DRC is regular to 100%. That notwithstanding, it should be considered that participants do not always provide the regular pronunciation (i.e., the most frequent answer) – they sometimes give responses that are quite different or even unexpected. This is an observation that DRC cannot account for, albeit to some extent through the increased involvement of the lexical route in nonword reading.

In principle, the proponents of DRC recognize the need to account for learning of GPC rules, to explain how children would acquire this set of rules. For one of their publications, Coltheart et al. (1993) devised a learning algorithm based on a database of spellings and pronunciations (2,897 words). Words were presented in random order; a word's spelling was given jointly with its phonetic transcription (Coltheart et al. 1993: 599). The algorithm was designed to deduce context-sensitive rules and position-specific rules from this input. Coltheart et al. (1993: 601) were pleased with the result: 78.17% of the words in the training set were read correctly. The GPC system made errors on exception words, but since exception words do not correspond to the regular conversion rules, errors on exception words are inherent in the model's design. Despite the good result of their work, Coltheart et al. (2001) abandoned this GPC learning approach, opting instead for fixed spelling-sound relationships. The authors had devised the

[3] The finding that nonwords take longer to read than real words.
[4] The finding that irregular words take longer to read than regular words.

algorithm in an effort to outperform the rival model PDP on its own database, i.e., the intention had been to show that DRC performed better on the same set of words. However, Coltheart et al. (2001: 216) state that the psychological plausibility of their learning approach had been criticized, and that there was no indication that this specially devised scheme did indeed reflect how children acquire conversion rules. Over two decades later, Pritchard et al. (2016) took up where Coltheart et al. (1993) had left off and developed a new rule-learning model named GPC-LM. The results were very similar to those obtained in 1993 (78.3% correct; Pritchard et al. 2016: 54-55), but a higher percentage of irregular words was read correctly. The question of plausibility remained, mainly because some choices made in training the model are arbitrary in nature: low-frequency correspondences are deleted, and the learning of single-letter, multi-letter and context rules happens in consecutive stages (Pritchard et al. 2016: 59-60). Still, the engagement with the topic of GPC rule acquisition is a step in the right direction, as predetermined rules, which are also perfectly regular, are an idealization that does not mirror human reading and its acquisition.

When it comes to reading aloud, orthography and phonology – letters and phonemes – are instrumental in the model; writing is translated into speech. Still, our understanding of the meaning behind the letter combinations ought to be considered as well. In DRC, there is a semantic system in the box-and-arrow diagram, linked to the orthographic as well as the phonological lexicon by means of arrows. In this box-and-arrow diagram (cf., Figure 2), these lexicons also have a direct connection, thus semantics appears to be an optional element, a secondary route, a bypass. In the computational version of DRC, semantics has not been considered, even though this failure to implement a semantic system is recognized as a weakness of the model (Coltheart 2005: 16). A reason for this omission is not given; apparently, the influence of the semantic system is perceived to be small, and perhaps not worth the effort at the time of formulating the basic processes of the model. Coltheart et al. (2001: 217) limit themselves to pointing to partial implementations and ideas realized in other studies; interestingly, one of them takes a semantic features approach based on WordNet (Miller 1995), which is also the basis for PDP's approach (cf., Section 3.2). Given the neglect of semantics in DRC, it is difficult to say how the system interacts with other parts of the model.

From this general overview of DRC, it seems that the model has its advantages as well as two main weaknesses. The model's architecture allows it to account for well-known effects found in word reading, such as lexicality, regularity, and frequency effects. DRC is incomplete insofar as its computational version completely lacks a semantic system, and the presence of predetermined conversion rules makes the model psychologically implausible, as it does not explain how these rules are acquired. Some efforts have been made to remedy the latter point of criticism.

3.2 Parallel distributed processing model of visual word recognition and pronunciation (PDP)

PDP was first put forward by Seidenberg & McClelland (1989) and explored further by Plaut et al. (1996); a semantic component was added to the model by Harm & Seidenberg (2004). Like DRC, it builds upon the Interactive Activation and Competition (IAC) model by McClelland and Rumelhart (1981) and Rumelhart and McClelland (1982). PDP uses distributed representations, which means that the model argues in favor of a distributed storage of information, as opposed to locally stored representations in a mental lexicon. The network consists of various structures or layers (input layer, intermediate or hidden layers, output layer) (Seidenberg & McClelland 1989: 526-527).

Most importantly, PDP is based on a connectionist network with "neuron-like *units* or *nodes* connected together so that a single unit has many links to other units" (Eysenck & Keane 2010: 23). These nodes are interconnected and may affect each other through excitation or inhibition; the unit takes the weighted sum of all of the input links, and produces a single output to another unit if the weighted sum exceeds some threshold value (2010: 23). A node receives input from many other nodes; the connections to these nodes do not all have equal importance or 'weight'. The weight of a connection is learned and may change over time; it depends on how often a pattern is presented to the model in training. This way, a certain input comes to be associated with a certain output. A connectionist model imitates the neural network of the human brain and is sometimes advocated as biologically plausible (e.g., Papadatou-Pastou 2011: 5).

In PDP, the representation of content is based on triples[5] (Seidenberg & McClelland 1989: 528). On the orthographic level, the word *make* would activate the units _MA, MAK, AKE, and KE_, whereby the underscore represents a word boundary. On the phonemic level, triples of phonetic features are activated, e.g., nasal, vowel, stop. The unit then activates possible letters based on a stored table (one table each for the word's beginning, middle, and end). There are 400 units on the orthographic level and 460 units on the phonemic level. 200 hidden units (or internal representation units) serve as intermediaries between orthographic input and phonemic output

[5] This scheme is based on prior work by Rumelhart & McClelland (1986), whose work is in turn based on Wickelgren (1969).

units; thousands of connections exist between the three layers. In contrast to DRC, PDP is built on learning; the model is trained on a corpus of 2,897 monosyllabic words that consist of three or more letters. How often these words are presented to the model depends on word frequency (530). After undergoing training, the model performed extremely well on words, with an error rate of only 2.7% (77 out of 2,897 words were read[6] incorrectly). Errors were either regularization errors (an exception word is pronounced like a regular word, e.g., *brooch* is pronounced /bru:tʃ/), incorrect vowels, or incorrect consonants (532). Performance on nonwords was not determined as a percentage by Seidenberg & McClelland (1989), but Besner et al. (1990: 434) report a performance of 51 % on nonwords and 59 % on PHs, based on McCann & Besner's (1987) set of stimuli. Seidenberg, McClelland & Kintsch (1990: 447-448) offer an explanation that builds on the amount of words in the training corpus, which is significantly smaller than the average vocabulary size of an adult (2,897 versus 30,000 words). They argue that if 30,000 words were used as input, the model would perform appropriately. As mentioned in the previous section, Coltheart et al. (1993) managed to outperform PDP on the same set of words with a learning model based on DRC.

Plaut et al. (1996) tackled the issue of poor nonword reading with a slightly revised implementation. They point out that a slot-based approach (letters being assigned to a particular slot, e.g., the first slot at the very beginning of a word) is incompatible with a connectionist architecture, as it presupposes position-specific knowledge instead of learned generalizations (Plaut et al. 1996: 64-65). Therefore, they abandon the triple-based approach and instead opt for grapheme and phoneme units; to avoid misplaced letters, an onset-nucleus-coda scheme was devised (65-66). This improved model was tested on 123 nonwords, and Plaut et al. (1996: 71) concluded that the model could "read pronounceable nonwords […] essentially as well as skilled readers". They give a percentage of 90.8 correct nonwords (Plaut et al. 1996: 70), but only after a few problematic words had been removed from the analysis.

In general, PDP does not entail the existence of a mental lexicon with entries for individual words (Seidenberg & McClelland 1989: 525); the model instead depends on the continuous adjustment of the connections between units. This is also how frequency and regularity effects are explained within the model, namely through "repeated adjustment of connection weights in the same direction" (535). The connections of frequent as well as regular words are adjusted more often, and they have greater influence on how the model performs.

[6] The model does not actually 'read' a word; it creates a phonetic representation. However, the term 'read' is typically used in the literature.

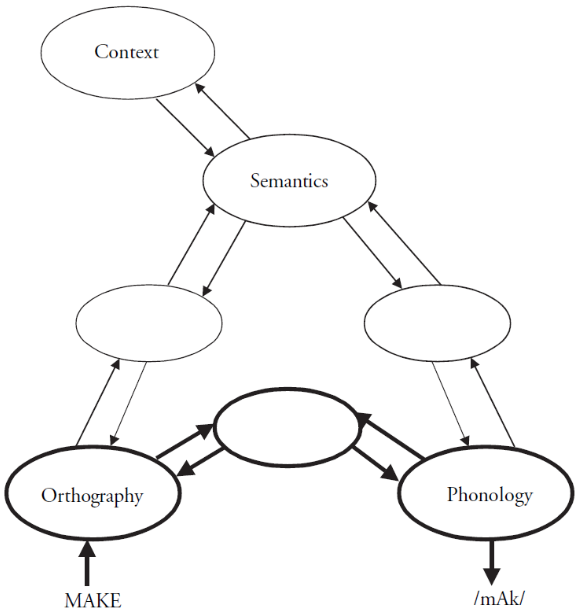


Figure 3 The orthography – phonology – semantics triangle of PDP
(reproduced by Coltheart 2005: 17 based on Seidenberg & McClelland 1989: 526)

PDP is often called a triangle model; Figure 3 shows this triangle that consists of orthography, phonology, and semantics. There are two routes: one direct route from orthography to phonology, and an indirect route from orthography to phonology via semantics / context. Insofar, PDP is a dual-route model, but unlike in DRC, this notion does not build on the separate processing of words and nonwords, but on a second route that operates via semantics. Seidenberg & McClelland (1989: 25) do not believe in separate routes for word and nonword processing: "A key feature of the model we propose is the assumption that there is a single, uniform procedure for computing a phonological representation from an orthographic representation that is applicable to irregular words and nonwords as well as regular words". In the original proposal, only the route from orthography to phonology was implemented (marked in bold in Figure 3). The second route via semantics was implemented by Harm & Seidenberg (2004); this paper focuses on reading and reading comprehension. Harm & Seidenberg (2004: 677) implemented a semantic system with semantic representations based on the online semantic database WordNet (Miller 1995). In this database, a number of relations between words are encoded, e.g., meronymy (part-whole). For the orthographic representation *dog*, examples of possible semantic features are [canine], [mammal], and [has_part_tail]. Harm & Seidenberg (2004: 677) used a training corpus of 6,103 words, for which 1,989 semantic features were generated. The number of semantic features per word ranged from 1 to 37. For a more detailed account of the WordNet approach, refer to Section 5.2. Given the neglect of semantics in competing models, the integrative approach of orthography – phonology – semantics in PDP is remarkable.

Harm & Seidenberg (2004: 667) state that their model tries to address "principles […] thought to underlie many aspects of perception and cognition", which means that they did not only focus on reading, but on the overall structure and plausibility. A major question addressed in their paper is the role of phonologically mediated semantic access. The model learns to access the semantic features directly from orthography (orth → sem) as well as via phonology (orth → phon → sem). The basic process is based on reading acquisition by children; they first learn to associate sounds with meaning, and orthography is later added when they learn to read. Accordingly, phon → sem was computed and trained first in PDP, and orthography was later added as an additional component. It is assumed that orth → phon → sem is the route preferred by children, but pressure to read fast leads to increased direct access without phonological mediation (Harm & Seidenberg 2004: 671). As already mentioned, PDP establishes and adjusts its connection weights in relation to the input received in training. The model learns through back-propagation; Eysenck & Keane (2010: 24) define back-propagation as "a mechanism allowing a network to learn to associate a particular input pattern with a given output pattern by comparing actual responses against correct ones". Harm & Seidenberg (2004: 672) compare this approach to a real-life situation in which a teacher provides the correct pronunciation to the students. PDP is based on the assumption that structures emerge gradually and are not predetermined.

In summarizing the main features of PDP, it should be emphasized that it is a connectionist model that uses distributed representations instead of a mental lexicon. The weights on the connections within the model constantly adapt in accordance with the input provided, which in this case is a training corpus of 2,897 words. Whereas the performance on real words was excellent, the performance on nonwords was somewhat poor. The absence of a mental lexicon complicates the explanation of phenomena such as the frequency effect, but Seidenberg & McClelland (1989) insist that these phenomena can be explained with reference to other factors inherent in PDP, e.g., the frequency-sensitive training of the model. To date, PDP is the only model for which a semantic system has been implemented. Harm & Seidenberg (2004) attempted to create a psychologically plausible acquisition process; the phonology-semantics pathway is trained first, orthography is added at a later stage. The notion of meaning is based on the activation of semantic features.

3.3 Connectionist dual-process plus model (CDP+)

CDP+ (Perry, Ziegler & Zorzi 2007) is based on CDP (Zorzi et al. 1998) as well as DRC. The goal of CDP+ was to formulate a new model that would build on the strengths of previous models, while improving on their weak spots. The creators adopted the lexical route of DRC with only minor adaptations, whereas a different non-lexical route that allows for learning was designed within CDP. Like PDP, the model possesses a (partly) connectionist architecture; like DRC, it is based on dual-route theory. These characteristics make CDP+ a kind of hybrid model.

The creators of CDP+ aimed to improve the poor nonword performance of connectionist models. To accomplish this, they take the non-lexical route (used for nonword reading) of CDP as a starting point; this system takes the form of "a fully parallel simple two layer associative network (i.e., a network without hidden units) trained to learn the mapping between orthography and phonology" (Perry, Ziegler & Zorzi 2007: 275). This network, called TLA for short, establishes the spelling-sound correspondences that are statistically most frequent, which basically means that it deducts (or learns) correspondence rules from input. To improve the existing TLA, a serialized left-to-right input of letters was chosen (as in DRC). A major reason for this choice is that serialized processing can explain serial effects[7] in word and nonword reading (279). In fact, Perry, Ziegler & Zorzi (2007: 289) state that CDP+ performs well with regard to serial/length and lexicality effects. Visual features are detected and parsed into letters; these letters are then assigned to onset, nucleus and coda slots (three slots in onset position, 1 vowel slot, and four slots in coda position). Slots that are not needed stay empty. These slots could be filled with 96 possible graphemes that are either simple or complex (281). The training corpus consisted of 7,383 letter combinations and 6,663 phoneme combinations; to simulate reading acquisition, pre-training was carried out with 115 grapheme-phoneme correspondences prior to the main training phase (281). During training, the orthographic frequency of the stimuli was taken into account (282). In the non-lexical route, a graphemic buffer (taken over from Houghton & Zorzi 2003) was added; this system parses the letters into locally represented graphemes. Its counterpart is a phonological output buffer located at the very end of the model – a decision system that makes the final choice of which phoneme is selected. The selection of phonemes is based on their respective activation level; the phoneme with the highest activation is selected, and partly, a certain threshold level must be exceeded (Perry, Ziegler & Zorzi 2007: 279).

[7] If processing goes from left to right and starts with the first letter, words with three letters are bound to be pronounced faster than words with six letters.

On close inspection, it becomes clear that CDP+ is based on an architecture that is both localist and connectionist. The lexical route presupposes the existence of lexical entries in a mental lexicon; there are orthographic as well as phonological entries. The non-lexical route, by contrast, is partly executed as a connectionist network. Graphemes and phonemes are represented locally, but the TLA Sublexical Network is a connectionist network that mediates between these localist representations. So while DRC and PDP opt for one of these main approaches only (i.e., they are either localist or connectionist in nature), CDP+ incorporates both of them. However, it would seem that localist representations are dominant, as the connectionist part appears to be integrated into a generally localist architecture. This is a very interesting approach that strives to capture the strengths of both sides: to build a learning model for nonwords, while readily accounting for frequency effects through the existence of a mental lexicon for real words. Figure 4 illustrates the architecture of CDP+.

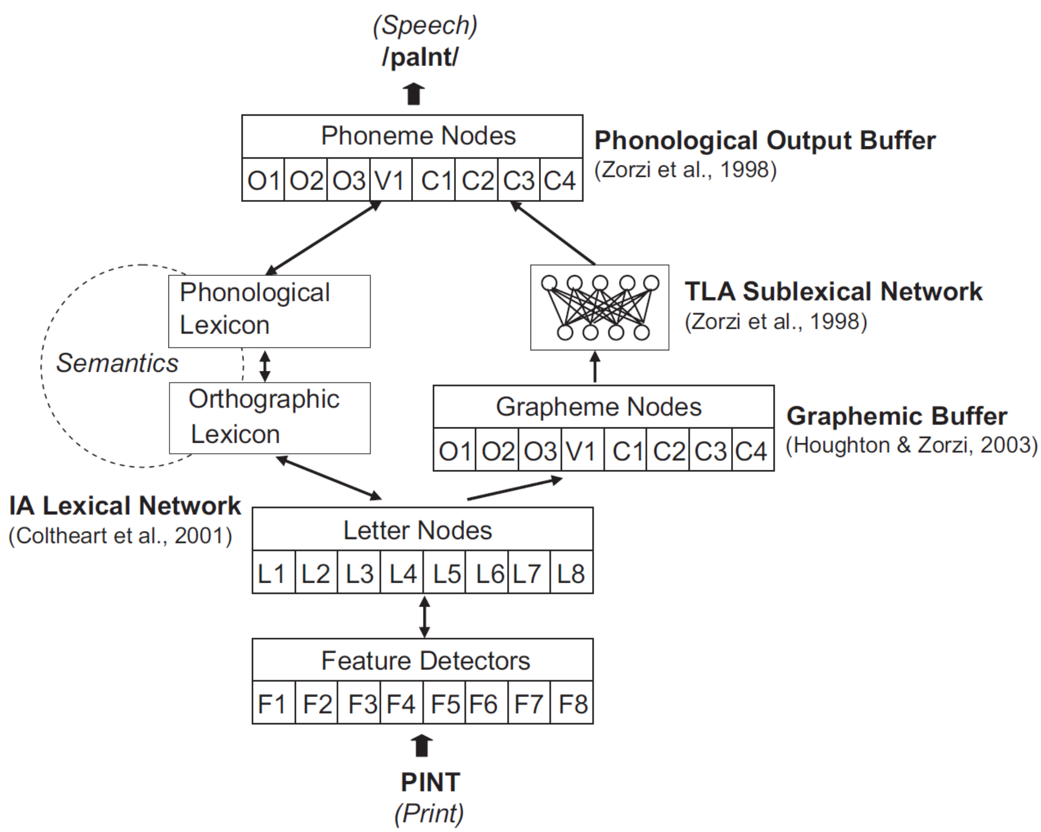


Figure 4 The architecture of CDP+ (Perry, Ziegler & Zorzi 2007: 280)

Just like DRC, and perhaps even more so, CDP+ lacks a semantic pathway. As the box-and-arrow diagram (Figure 4) shows, semantics is hypothesized to exist somewhere along the lexical route, but Perry, Ziegler & Zorzi (2007: 278) made a conscious choice to exclude it from their model. They give three reasons for this decision: the controversial role of semantics in reading aloud in general, the questionable explanation of how lexical decision works in connectionist models (as formerly presented by other researchers), and simpler testing of nonword reading with a reduced lexical route. Overall, Perry, Ziegler & Zorzi (2007: 303) believe that the contribution of semantics in word naming is very limited. While it is understandable that semantics would be left out at first for the sake of simplicity, it should also be clear that semantics needs to be added at a later stage in order to render a model comprehensive.

Following the creation of CDP+, the model's performance on words and nonwords was evaluated (Perry, Ziegler & Zorzi 2007: 285). Out of the 7,383 words included in the training database, 98.67% were read correctly. Out of 592 nonwords, 93.75% were read correctly, but it should be pointed out that the authors adopted a fairly lenient measure of correctness – a fact that they even acknowledge themselves. Only responses that showed illegal grapheme-phoneme correspondences were counted as errors. Pritchard et al. (2012: 1270) point out that many responses that were counted as correct are unlikely to be given by actual participants, e.g., /bʌnθ/ is an unlikely response to *bonth*. Hence, it should be kept in mind that the error rate of CDP+ could be much higher if a different measure of correctness was adopted. A strength of the model is the variety of nonword responses that it produces. As pointed out by Perry, Ziegler & Zorzi (2007: 277), the model "delivers not only the most common mapping of a grapheme but also less common mappings". This is a strength precisely because it matches human data; participants do not always produce the 'top answer' in response to a nonword. According to Pritchard et al. (2012), who collected nonword responses from 45 Australian students, participants produce quite a number of unexpected responses. Frequent responses are lexicalization errors (a real word response to a nonword), dropped or added phonemes, the choice of a different vowel or consonant, and an s/z difference in coda position (cf., Pritchard et al. 2012: 1276). By design, CDP+ can explain such varying responses.

As mentioned in this section, CDP+ is a hybrid model that combines a dual-route architecture and localist representations with a connectionist system for nonword reading. A strength of the model is the variety of nonword responses that it produces (it better matches findings from human data because its responses are not entirely regular); a weakness is the complete absence of a semantic route. It should be pointed out that the evaluation of nonword reading depends greatly on the parameters used to measure performance. Taking all factors into account, it appears that CDP+ is a promising computational model. Unlike DRC, it allows for training of grapheme-phoneme correspondences; contrary to PDP, it builds on the assumption of lexical entries and can thus explain many well-known effects found in the area of reading aloud.

4. The pseudohomophone effect

The question if PHs are processed differently from pseudowords has received much attention in the literature. The basic assumption is that the naming times of PHs differ from those of pseudowords because their sound pattern is known to the reader. Moreover, PHs have meaning, but this aspect is typically ignored in psychological experiments of this kind. Both pseudowords and PHs are regarded as nonexistent words, even though words such as *focks* could trigger associations connected with the base-word *fox*. Instead of semantics, studies focus on orthographic and phonological aspects of words, and aim to assess how certain characteristics of word types cause the reaction times of participants to change. Due to their special nature, PHs have been used quite extensively for the purpose of studying reading aloud, and any interesting findings have been said to support or contradict common views or well-known models. On the whole, researchers have employed creative methods in their investigation of the foundations of reading aloud. Since a textbook approach to these methods is rarely encountered in the literature, the present section will first present an overview of the most frequent methods used to investigate the processing of PHs. Subsequently, the main findings in this research area will be discussed.

The following **methods** have conventionally been used in investigating the processing of PHs:

- <u>naming task</u> (collecting naming times)
  In this field, naming refers to the reading aloud of an item that is displayed on a screen. The time that elapses from the presentation of a word to the onset of articulation is measured. Naming times are often referred to as naming latencies. A word should be named by a participant as fast as possible. The purpose is to compare the naming latencies of PHs, real words and/or nonwords to see if there are faster or slower naming times for these word groups by item and/or participant.
- <u>delayed naming task</u>
  Participants are asked to pronounce the words a second time and are allowed more time (~1 second) to prepare a response. This method serves to determine if naming speed is related to ease of articulation (e.g., Taft & Russell 1992; Grainger, Spinelli & Ferrand 2000). In delayed naming, the summed naming times for nonwords and PHs should be roughly equal; if PHs are pronounced faster, it is likely that they were just easier to

pronounce because they possess familiar structures (and not because a lexical entry was activated).

- establishing the frequencies of the base-words to check for a base-word frequency effect
  The word frequency effect based on Morton's (1969) logogen model stipulates that words which are accessed frequently have a lower activation threshold than words which are accessed infrequently. The presence of a base-word frequency effect is generally seen as an indication that a word's entry in a presumed mental lexicon has been activated.
- lexical decision task
  Participants are asked if a letter combination represents an existing word. They answer with yes or no by pressing a button on a response box. Response times are collected and interpreted. The experiment design stipulates that PHs are not real words: 'no' is the correct answer in response to a PH. It is likely that subjects will have more difficulty in deciding if a PH is a word because phonology interferes with their decision.
- phonological lexical decision task
  Participants are asked if a letter combination sounds like an existing word (e.g., McCann, Besner & Davelaar 1988; Grainger, Spinnelli & Ferrand 2000). They answer with yes or no by pressing a button on a response box. This task represents a counterpart to the lexical decision task.

Using the methods outlined above, the processing of PHs has been studied extensively with different outcomes. A common finding in studies of reading aloud single words is the **'pseudohomophone effect'**. This effect is said to occur if the naming latencies of PHs are different from those of (similar) nonwords. The direction of this effect can be twofold: a PH advantage, meaning that PHs are pronounced faster than nonwords, or a PH disadvantage, meaning that PHs take longer to pronounce than nonwords.

In a mixed list design (in which PHs are mixed with nonwords), the most common finding in the literature is a PH advantage with no apparent base-word frequency effect (McCann & Besner 1987; Herdman, LeFevre & Greenham 1996; Grainger, Spinelli & Ferrand 2000). This means that PHs are named faster, but the frequency of their base-words does not seem to play a role. There are 2 main explanations for these faster naming times: ease of articulation (McCann & Besner 1987; Taft & Russell 1992; Herdman, LeFevre & Greenham 1996; Seidenberg et al. 1996) and orthographic similarity to real words (a point first made by Martin 1982). Ease of articulation means that PHs are pronounced more easily and rapidly because people have experience in uttering these words. As Herdman, LeFevre & Greenham (1996: 1057) put it, faster naming reflects "the ease with which phonology is mapped onto articulatory units", or the access of "whole word representations in the phonological output lexicon" (McCann & Besner 1987: 14). To put it simply, PHs are pronounced faster because their pronunciations is identical to that of real words, and skilled readers have practice in pronouncing these words. With regard to orthographic similarity, Martin (1982) investigated this idea using a novel research design. In her lexical decision experiment, she used a visual control: a nonword that differed from the PH by just one phoneme. This approach produced nonwords that were visually very similar to PHs (e.g., *werd* / *serd*); in addition, Martin (1982) used nonwords that were visually distant to compare the results of the two groups. Interestingly, Martin (1982: 401) found no difference in the results obtained on PHs and the visually similar nonwords, but only between PHs and visually distant nonwords. She therefore argues that visual similarity to real words has a stronger influence on nonword reading than phonological similarity (Martin 1982: 395).

Researchers' views on lexical decision do not vary greatly. Overall, they agree that the lexical decision task requires a different kind of processing than naming tasks. It seems that a disadvantage of PHs is generally seen as the dominant finding in lexical decision (cf., Ziegler, Jacobs & Klüppel 2001: 547). Stone & Orden (1993) argue that words which are more similar to real words take longer to reject; this would mean that the response to PHs in lexical decision is slower (even though they are read aloud faster in the naming task). Seidenberg et al. (1996: 57) believe that semantic activation takes place in lexical decision (and thus influences reaction times), but not in naming tasks. Besner & Davelaar (1983) likewise found a PH disadvantage, despite their efforts to control for visual similarity. As a result, they contradict Martin (1982) and argue in favor of a PH effect (i.e., PHs are processed differently from real words) that they attribute to the access of phonological information during the task (Besner & Davelaar 1983: 303). Their explanation is that their own PHs were largely homophonic (meaning that they had two base-words, e.g., *woar* has the base-words *war* and *wore*), whereas Martin's stimuli were not. McCann, Besner & Davelaar (1988) extended their study (in which they found a PH disadvantage and no base-word frequency effect) by adding a phonological lexical decision task, i.e., subjects were required to state if the PH in question *sounded* like a real word. In this task, the authors did find a base-word frequency effect, as did Grainger, Spinelli and Ferrand (2000). McCann, Besner & Davelaar (1988: 703-704) believe that naming and lexical decision rely principally on the orthographic lexicon, with phonology playing a small role. The phonological lexical decision task foregrounds phonology, which is why the frequency effect appears. It thus seems that lexical decision and phonological lexical decision necessitate the access of different kinds of information. At this point, it should be noted that Taft & Russell (1992) contradict the dominant view regarding reaction times in lexical decision; they found faster reaction times to PHs (60) and suggest that their participants took longer to respond to nonwords because they kept reanalyzing them in an effort to find a real-word match.

It has frequently been argued (Marmurek & Kwantes 1996; Grainger, Spinelli & Ferrand 2000; Borowsky, Owen & Masson 2002) that the direction of a PH effect varies according to stimulus presentation. Overall, stimuli may be presented in mixed lists or pure blocks. In mixed lists, types of stimuli are displayed in a mixed order, i.e., nonword – PH – nonword – PH. In pure blocks, all stimuli of one type are displayed in one block, i.e., nonword – nonword – nonword. In the latter scenario, PHs may be presented at the beginning or at the end of a task. Borowsky, Owen & Masson (2002) tested multiple scenarios and found a base-word frequency effect as well as a PH disadvantage for three of four sets of stimuli, provided that PHs were presented first in pure blocks. Marmurek & Kwantes (1996) proffer that in pure lists, access of lexical entries is facilitated as opposed to mixed lists. Reynolds & Besner (2005) support the view of Borowsky, Owen & Masson (2002) and present a novel explanation of this phenomenon. Specifically, Reynolds & Besner (2005: 633) proprose that PHs are read aloud by reference to "general word knowledge" – in a mixed-list naming task, orthographically or phonologically similar lexical entries are activated, but PHs activate their base-words more strongly because an exact match is found, which results in a PH advantage. However, a base-word frequency effect does not occur, as the base-word's contribution is blurred by the activation of co-activated entries. But if PHs are read aloud prior to nonwords in a pure list, the corresponding base-word is activated through a "specific activation strategy" (633), meaning that a very specific search process for the target word is performed. This specific search takes longer to perform, resulting in a PH disadvantage. This specific search process also leads to a base-word frequency effect, given that the base-word is activated more strongly and directly. In conclusion, it would appear that stimulus presentation in mixed lists and pure lists leads to the application of different strategies by the reader, which manifests itself in different outcomes (naming latencies, base-word frequency effect).

Grainger, Spinelli & Ferrand (2000) provide a different interpretation of the results obtained in mixed lists versus pure blocks. In a pure-block design, Grainger, Spinelli & Ferrand (2000: 92) found that real French words and PHs behaved similarly as both showed a significant effect of neighborhood size and frequency[8]. Changing the conditions from pure blocks to mixed lists of PHs and nonwords, a base-word frequency effect could not be attested (93); in the phonological lexical decision task, the effect reappeared (95). Hence, Grainger, Spinelli & Ferrand (2000: 94) propose that PHs are processed similarly to words when presented in pure blocks (and the participants are aware of the nature of the stimuli), whereas in mixed lists, they are processed like nonwords. In their interpretation of the observed effects, the authors turn to dual-route models of naming, stating that the lexical route is dominant in pure blocks, whereas the sublexical route is dominant in mixed list presentation (96). What this interpretation cannot explain is the PH advantage in mixed lists (97). Moreover, real words and PHs were not tested in a mixed list design, and therefore a direct comparison between pure blocks and mixed lists is not possible. Such a mixed list design using real words and PHs will be the subject of study of this thesis (cf., Section 6).

As regards the design of nonword experiments, it is interesting to note that the number of participants tends to vary greatly. There seems to be no consensus on what constitutes a representative sample or even a discussion about this aspect of the research – the number of participants ranges from 16 (Marmurek & Kwantes 1996, Experiment 1) to 120 (Borowsky, Owen & Masson 2002, Experiment 1). In part, the total number may depend on the sub-groups needed if different tasks are performed by different groups. The majority of researchers opts for a number between 20 and 48; usually, university students participate for course credit and thus the number also depends on availability. In experiments of this nature, it is common practice to perform a by-subjects as well as by-items analysis, as this will reveal differences between participants in naming latencies or reaction times. Taft & Russell (1992: 65) observed that the participants in their experiment could be divided into slow and fast readers according to naming latencies. A base-word frequency effect could only be attested for the slow group, and Taft & Russell (1992: 64) hypothesize that the longer naming times in this group allowed the lexical route to play a bigger part in reading aloud. Marmurek & Kwantes (1996: 709) make a valid point in stating that participants' awareness of the nature of the stimuli factors into their approach to the task; if they know that the PHs are unusually spelled versions of a real word, the naming task is extended by another dimension – that of identifying the PH's base-word. Consequently, the researcher needs to make a conscious choice in the research design and define whether or not to instruct the participants as to the nature of the stimuli.

[8] It should be mentioned that the PH stimuli of Grainger, Spinelli & Ferrand (2000) were created from their base-words by changing just one letter and were thus visually very similar to real words.

Another reason for the differing results obtained in PH experiments lies in the nature of the stimuli; researchers (e.g., Borowsky & Masson 1999) have pointed out that only PHs that may be recognized as such should be used as stimuli. Indeed, appendices to frequently cited studies list PHs that do not readily relate to the intended target word, such as *bawk* or *fers*. If the connection between the PH and its base-word cannot be established, this missing link is bound to influence the outcome of the study. Therefore, it seems useful to corroborate participants' recognition of the word as a PH by means of a lexical decision task, and exclude items that were not recognized from the analysis.

Table 1 The pseudohomophone effect: literature review / findings

Method of stimulus presentation	Findings
mixed lists	PH advantage, no base-word frequency effect (McCann & Besner 1987; Herdman, LeFevre & Greenham 1996; Grainger, Spinelli & Ferrand 2000 Experiment 2) PH advantage, base-word frequency effect (Taft & Russell 1992, but only for the slow group)
pure blocks	PHs only: base-word frequency effect (Grainger, Spinelli & Ferrand 2000, Experiment 1, compared to real words) PHs first: PH disadvantage, base-word frequency effect (Borowsky, Owen & Masson 2002; Reynolds & Besner 2005)
lexical decision task	PH disadvantage (Besner & Davelaar 1983; McCann, Besner & Davelaar 1988 Experiment 1, no base-word frequency effect; Stone & Orden 1993; Seidenberg et al. 1996) PH advantage (Taft & Russell 1992)
phonological lexical decision task	base-word frequency effect (McCann, Besner & Davelaar 1988 Experiment 2; Grainger, Spinelli & Ferrand 2000)

In conclusion, it may be stated that the common finding in PH naming varies according to the order in which the stimuli are presented. In mixed lists, the standard finding is a PH advantage over nonwords with no apparent base-word frequency effect. In pure blocks, it is more likely to find a PH disadvantage accompanied by a base-word frequency effect. In lexical decision, a PH disadvantage (PHs take longer to dismiss than real words) has come to be seen as preferred, and a base-word frequency effect tends to be found in phonological lexical decision. In general, it is accepted that the task that is asked of the participants has a major influence on the results. Pure lists may produce different results than mixed lists, simply because they lead participants to apply a different strategy in reading aloud.

4.1 The PH effect in DRC

After performing a literature review, Coltheart et al. (2001: 225) conclude that a PH advantage in naming is caused by orthographic similarity to real words, by a base-word frequency effect in slow naming (based on Taft & Russell 1992), as well as by presenting the stimuli in pure lists (based on Marmurek & Kwantes 1996). The common explanation that the effect results from ease of articulation is not mentioned. Coltheart et al. (2001: 221) carried out a test with 80 PHs and 80 matched nonwords. They state that their model produces a PH advantage, just as would be expected from human participant data. This advantage appears to be based on the involvement of the lexical route in PH reading, as the effect disappears once the lexical route is disabled. Coltheart et al. (2001: 221) also used Taft & Russell's (1992) and McCann & Besner's (1987) stimuli and report that DRC manages to mirror the effects that were found in these papers. However, they fail to acknowledge that their explanation disregards the most frequently cited argument in the literature. If the involvement of the lexical route is responsible for faster naming, this stands in contrast to ease of articulation as the primary source of the effect.

In lexical decision, DRC predicts that PHs take longer to reject than pseudowords. This difference is attributed to a longer processing route: a PH is first parsed by the non-lexical route (according to the GPC rules) until a pronunciation has been computed at the phoneme level; this pronunciation then activates the phonological lexicon, and in turn the orthographic lexicon, in a bottom-up manner (Coltheart et al. 2001: 230-231). By following this path in the box-and-arrow diagram (Figure 2), this argument becomes clearer. Accordingly, Coltheart et al. (2001: 230) hypothesize that PHs will take longer to reject than nonwords, as the participant will take longer to decide due to the 'look-up procedure' that is performed for PHs. Simply put, subjects need more time to answer the question 'Is the stimulus a real word or not?'. In addition, Coltheart et al. (2001: 231) make a distinction between PHs that are orthographically very similar to their base-words (e.g., *koat* – *coat*) and words that are less similar (e.g., *kote* – *coat*).

Building on prior work, the authors believe that PHs that show greater similarity will receive stronger activation from the lexical route than their less similar counterparts, which influences their processing. In a lexical decision experiment, this means that *koat* takes longer to reject than *kote*. Overall, it may be concluded that DRC is in line with the common explanations of lexical decision performance found in the literature.

4.2 The PH effect in PDP

Seidenberg & McClelland (1989: 555) agree with the overall findings for PHs: lexical decision ought to take longer than for nonwords, whereas naming should be more swift. They also realize that the absence of a mental lexicon with lexical entries is likely to be a problem in accounting for differences between PH and nonword reading. Still, they state that their model is able to simulate both effects. In the case of PH naming, Seidenberg & McClelland (1989: 555) believe that PHs are named faster only if they resemble real words more closely than nonwords do (visual similarity as pointed out by Martin 1982). For PHs that are less word-like (which is inferred from higher error rates of the model), PDP predicts that there will be no difference in word naming. In the case of lexical decision, Seidenberg & McClelland (1989: 550) assume that subjects differentiate between words on three levels. If the decision is between real words and unpronounceable nonwords, orthographic information suffices to make a decision (on the basis of illegal letter combinations). If the decision is between real words and pronounceable nonwords, phonological information is needed (to assess if the word sounds familiar). In the case of PHs, semantic information is key: Does the stimulus trigger any semantic features? These three types of information are activated at the same time. In 1989, this was still a theoretical assumption, as there was no implemented semantic pathway. A possible 4^{th} dimension that might come into play is contextual information.

Harm & Seidenberg (2004) are able to make more detailed statements on PH processing. After running a set of PHs through PDP, they conclude that “the bulk of the activation of semantics is done via the phonological pathway, as would be expected” (Harm & Seidenberg 2004: 706), but a part of the activation also stems from the orthographic pathway. This is typically the case with PHs which are very similar to their base-words (e.g., *ghoast* – *ghost*) and have few orthographic neighbors (Harm & Seidenberg 2004: 706-707).

4.3 The PH effect in CDP+

To investigate the performance of CDP+ on PHs, Perry, Ziegler & Zorzi (2007: 293) used the stimuli of McCann & Besner (1987). The authors found a PH advantage for these stimuli, as well as a weak correlation between naming latencies and word frequency. Referring to Reynolds & Besner (2005), the authors adopt the view that a base-word frequency effect only occurs if the stimuli are presented in pure blocks (it depends on reading strategy and is therefore task-specific). To simulate a pure-block condition, one of the parameters of the model was altered (the phoneme activation criterion was raised from 0.67 to 0.73); this alteration results in an increased base-word frequency effect. This is the case because the threshold for activating a phoneme is higher, and possibly allows the model time to access lexical entries. Changing this parameter also leads to a longer naming time of PHs: the result is a PH disadvantage in pure lists. Perry, Ziegler & Zorzi (2007: 293) see this parameter change as very efficient: only one parameter needs to be changed in order to account for the different performance on pure lists. To sum up, CDP+ appears to be in line with a PH advantage for mixed lists and a PH disadvantage for pure blocks, whereby a clear frequency effect only occurs in the latter condition.

With regard to lexical decision, Perry, Ziegler & Zorzi (2007) do not make any claims. This may seem a little surprising at first, but considering that lexical decision is actually unrelated to the process of reading words aloud, it is understandable that they would not make any statements. Any predictions made could not be tested within a model that is designed for naming only. In general, though, Perry, Ziegler & Zorzi (2012: 292) state that their model would perform similarly on lexical decision as DRC, given their nearly identical lexical route.

5. The mental lexicon and its semantic component

As pointed out in Section 3, semantics is rarely implemented in models of reading aloud. A possible reason for the neglect of semantics is the rather elusive nature of meaning in general. It is therefore useful to take a closer look at meaning, i.e., at the mental lexicon and its organization. The relations between words (e.g., synonymy, antonymy, hyponymy) have been the subject of detailed linguistic study. For the question of what meaning actually is, however, it is necessary to turn to philosophical approaches. This section will conclude with the presentation of possible influences of semantics on the process of reading aloud.

5.1 The organization of words in the brain

As observed by Aitchison (2012: 16), “humans know tens of thousands of words, most of which they can locate in a fraction of a second”. It is commonly believed that words are stored in a mental lexicon, but as mentioned in Sections 3 and 4, not all researchers believe in this concept. Eysenck & Keane (2010: 340) define a mental lexicon as “a store of detailed information about words, including orthographic, phonological, semantic, and syntactic knowledge”. It is important to understand that the lexicon consists of all of these components, and that these components need not all be accessed, or be accessed at the same time. What kind of information is accessed is often unconscious and strongly task-dependent. Thus, lexical access is a fairly broad term if it is not specified further. At this point in time, we can only hypothesize how the mental lexicon is organized; we do not know if the four components are stored together, or if they are stored in a distributed manner, scattered across the brain. It is further conceivable that the four components themselves may be further split into various kinds of information.

In her work *Words in the mind: an introduction to the mental lexicon*, Aitchison (2012) focuses on the semantic component of the mental lexicon. The basic idea underlying this work is that our knowledge of words is organized in semantic networks. Ahlsén (2006: 83) defines semantic fields or networks as “groupings of words according to semantic similarity, or contiguity (co-occurrence), relations”. Aitchison (2012: 102) identifies the most frequently named relations between words as co-ordination (word pairs like *salt* / *pepper* or opposites such as *black* / *white*), collocations (e.g., *salt* / *water*), superordination (*color* / *red*), and synonymy (*starving* / *hungry*). These relations result in a web of semantic fields. However, these relations are by no means all possible semantic relations; Fromkin, Rodman & Hyams (2003: 179-184) list homonyms, homographs, heteronyms, synonyms, antonyms, hyponyms, metonyms, and retronyms. Whether these relations are encoded in the brain in some way is not directly observable, but may be inferred from linguistic ‘evidence’ such as slips of the tongue, priming, or linguistic disorders (e.g., anomia, semantic dementia) (176). An example of a slip of the tongue would be “bridge of the neck” instead of “bridge of the nose” (176), whereby the target word and the actually uttered word are both body parts that start with the same phoneme. A similar case is the tip of the tongue phenomenon, which was first investigated in more detail by Brown & McNeill in 1966. They supplied a definition of a fairly infrequent word to groups of students, and asked them to identify characteristics of the target word if they were just on the verge of identifying it. Brown & McNeill (1966: 329) found that the initial letter of the word was guessed correctly in 57 % of cases, and students also guessed well in the case of the number of syllables as well as the stressed syllable of the target word (330). A particularly interesting finding was that the initial letters of a word seemed to receive the highest activation (they were often guessed correctly), followed by the final letters of the word. The words' middle letters, however, rarely matched the target words (330). Aitchison (2012: 158) calls this the "bathtub effect": The beginning and end of a word are above the water (like the head and the feet of a person lying in a bathtub), but the middle of the word stays below the surface. This middle part can be filled with many different letters, e.g., an_ _ _ _ _ _dote would be filled with either *anecdote* or *antidote* (159). It follows that word substitutions are not random, but share some kind of semantic or phonological property. This finding indicates that words are part of a semantic network; words help define other words, which leads to a sort of 'interrelatedness'.

Many researchers have tried to uncover a basic structure underlying word categorization. Well-known approaches are the semantic features approach and prototype theory. Fromkin, Rodman & Hyams (2003: 177) define semantic features as "a formal or notational device that indicates the presence or absence of semantic properties by pluses or minuses". If a certain feature applies to a word, this is expressed by a plus sign. If the feature does not apply, this is expressed by a minus. Table 2 illustrates how semantic features apply to the words *man*, *woman*, *boy*, and *girl*.

Table 2 The semantic features approach
(adapted from Fromkin, Rodman & Hyams 2003: 176)

man	woman	boy	girl
+ human	+ human	+ human	+ human
+ male	- male	+ male	- male
+ adult	+ adult	- adult	- adult

This approach makes it possible to quickly identify the main similarities and differences between two words, and it highlights their core meaning. For many words, however, it is almost impossible to come up with such a core meaning, especially in the case of abstract concepts such as *love*. It is clear that a mere list of features does not fully illustrate word meaning, and as can be seen in Table 2, many words come to be described not by what they are, but by what they are not (in this case, 'male' or 'adult'). Moreover, associative meaning (connotations, culturally shared or learned beliefs) form another aspect of word meaning and may be activated along with a word. The core properties of a word are hard to select; sometimes, it is difficult to determine into which category a word should belong. Precisely this 'fuzziness' is the central idea of prototype theory (e.g., Rosch 1975). In prototype theory, a member of a class may be defined as an ideal fit or prototype, or as less prototypical, whereby there is a continuum of prototypicality. Figure 5 illustrates this concept with the example of 'birdiness' (taken from Aitchison (2012: 69): a robin is a prototypical bird (at least to a European or an American), an owl is already less prototypical, and a penguin or an ostrich would be considered as barely belonging to the category 'bird'.

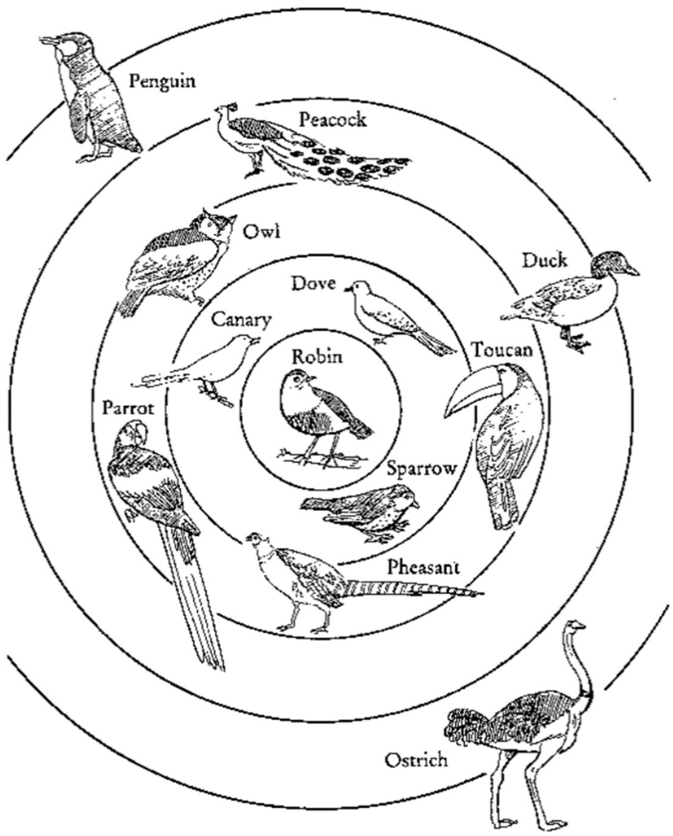


Figure 5 Prototype theory (an illustration from Aitchison 2012: 69)

Figure 5 is actually based on an experiment carried out by Rosch (1975). Participants were asked to complete a questionnaire in which they ranked words according to their memberships in categories such as bird, vegetable, or furniture. The rankings were surprisingly consistent across the participants. It could therefore be argued that human beings tend to organize their knowledge in categories, and that this tendency is visible in language as well. An advantage of the prototype approach is that it allows for gradience (as opposed to the semantic features approach); there are typical as well as less typical members of a category. Members of the same category may share a set of semantic features, or trigger similar associations.

An interesting question in the context of this paper is how words are selected in a regular conversation. In order to utter a word, this word has to be selected first from a range of possible candidates. This means that semantics precedes the selection of phonemes, human beings first need to identify the proper word or word sequence (e.g., formulaic expressions) that most closely matches what they are trying to express and also fits the intended syntactic structure. According to Aitchison (2012: 245), a semantic field is selected and narrowed down; subsequently, activation spreads to the phonological lexicon, where possible beginnings and endings of words are activated, as well as within the semantic lexicon. This activation on the level of sound and meaning is illustrated in Figure 6 below.

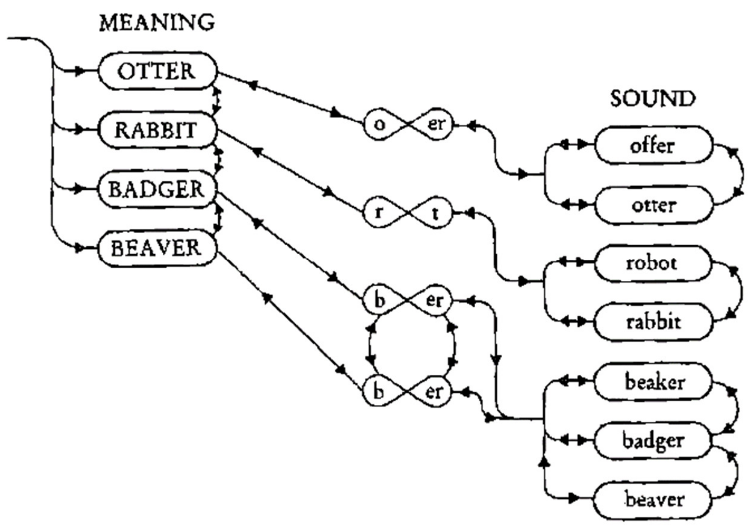


Figure 6 An interactive activation model (an illustration from Aitchison 2012: 245)

The relevant links get a higher level of excitation; conscious attention of the speaker results in the selection of the appropriate word. This interplay between semantics and phonology may explain why a wrong word choice often results in the selection of a competitor that is related in both domains (as in the tip of the tongue example *nose* / *neck*). The concept described here is the basis of interactive activation models. One such model is the Interactive Activation Model by McClelland & Rumelhart (1981) mentioned in Section 3. In this model, top-down and bottom-up processes interact, leading to an interplay of word level, letter level, and feature level in visual word recognition, i.e., reading. This interplay can explain why letters are identified more easily in words than in isolation or in nonwords because the activation of whole-word entries leads to increased overall activation that facilitates detection (Eysenck & Keane 2010: 338), e.g., the identification of the letter R in CARD as opposed to CRDA.

Whereas semantics precedes phonology in regular speech, this sequence might not necessarily apply to reading aloud. In the case of real words, it might well be the case, especially if the word is very frequent. It is also conceivable that we make meaning only after a word has been read, and so semantics would be activated subsequently to phonology. At times, we might not recognize the meaning of a sentence at all, and will need to re-read certain passages. In the case of pseudowords, the first intuition would be that they do not trigger word meaning at all, except perhaps for the detection of word class, as in the jabberwocky poem. Still, word meaning might come into play if a pseudoword is visually similar to an existing word. Considering that the beginning and end of a word is key in word recognition, a whole range of words could potentially be activated in nonword reading, even if this activation is only weak. This feeble activation could lead a reader to associate a nonword with semantic properties. The same is the case in the reading of PHs: visual similarity could lead to the activation of real word entries in the mental lexicon. This activation might lead to regularization errors (i.e., a nonword being pronounced like a real word) or to the computation of phonology by means of analogy. But much more often, the phonology of pseudowords will be computed on the basis of a rule system (cf., Section 3), and if this sound pattern matches that of an existing word, this match could hypothetically lead to the activation of the word's meaning.

5.2 Approaches to the conceptualization of meaning

A fundamental question underlying research on semantics in reading aloud is how meaning can be conceptualized for this purpose. What is 'meaning', and what information about words is stored in our brain? While linguistics is concerned with exploring how words are related to each other, it does not attempt to explain what meaning actually is. At its most basic, the question 'What is meaning?' is a philosophical one. This sub-section will therefore start out with Saussure's (1986) concept of the 'linguistic sign' and Frege's (1948) approach to reference, sense, and associated conception. Subsequently, Miller's (1995) semantic database WordNet will be scrutinized. Saussure's approach has been highly influential in linguistics in general; Miller's approach has been implemented in two computational models (PDP – Harm & Seidenberg 2004; DRC – Pritchard et al. 2016). Overall, it cannot be denied that meaning is a more elusive concept than orthography or phonology, which makes it hard to address semantics from a scientific point of view.

In researching language, Saussure makes a distinction between mental concepts and physical realizations of language. His concept of meaning centers around the 'linguistic sign': "the combination of a concept and a sound pattern" (Saussure 1986: 67). Figure 7 illustrates the notion of the linguistic sign.

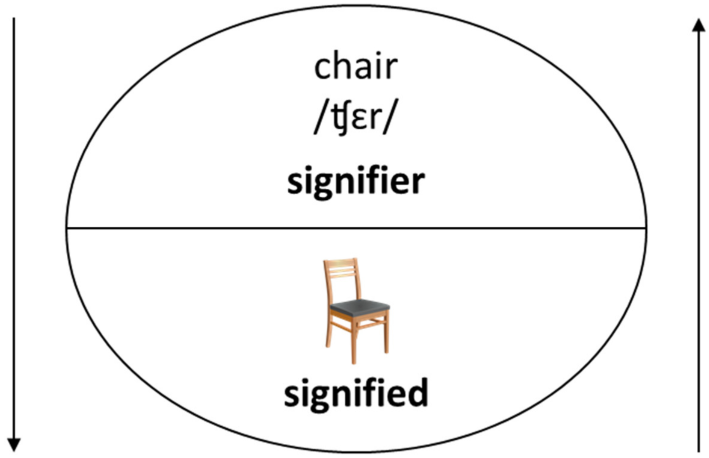


Figure 7 The linguistic sign (adapted from Saussure 1986: 67)

The relation between the concept and the sound pattern (or signal) is arbitrary, except for special cases such as onomatopoeic words or interjections (67-69). As regards the sound pattern or signal, Saussure (1986: 69-70) emphasizes that it is fixed in time and can only be measured in a linear fashion. Furthermore, he observes that "[t]he sound of a word is not in itself important, but the phonetic contrasts which allow us to distinguish that word from any other" (Saussure 1986: 116) – only phonology is part of the linguistic sign as it exists in the mind, but phonetics and individual letters are vital in accessing the sign. Similarly, the conceptual aspect is defined through neighboring concepts; their meaning helps delimit the concept at hand, for example the existence of *mutton* restricts the meaning of the word *sheep*. Saussure (1986: 114) states that "the value of any given word is determined by what other words there are in that particular area of the vocabulary. […] No word has a value that can be identified independently of what else there is in its vicinity". When contemplating the concept of the signifier and the signified, one problem presents itself. Ideally, a signifier would refer to a prototypical object, but in fact, signs often refer to concepts or ideas whose imageability is low. These concepts cannot always be pictured in the mind, and thus it is glaringly evident that Figure 7 is only a simplification of how meaning is stored in the brain.

In his well-known work *Sense and reference*, Frege (1948) expresses a different view of the sign. For Frege (1948: 210), a sign refers to "a proper name […] whose designation can also consist of several words or other signs". In addition, Frege (1948) establishes three main distinctions: reference, sense, and the associated conception. In its most basic sense, 'referent' refers to “an object perceivable by the senses” (212). In contrast, ‘sense’ refers to the designation under which this referent is known (*morning star* and *evening star* both refer to the planet Venus), and ‘conception’ refers to the individual associations that a person has with the object or concept in question (212). The main contrast to Saussure’s work is that Frege includes the actual object in the outside world in his triangle of meaning; thus, his conceptualization is not purely psychological. Moreover, Frege’s (1948) notion of the sense draws attention to the fact that a referent can be known under different names (e.g., in the case of an author’s pseudonym), and that a sign can stand for different referents, caused by phenomena such as polysemy and homophony. By establishing the associated conception, Frege (1948) makes it clear that signs can mean different things to different people, even if the sense is the same. This concept is similar to the term ‘connotation’, but a connotation is often shared by a group of speakers (and may therefore differ between cultures), whereas the individual associations can vary greatly.

A more scientific approach that focuses on how meaning is organized is WordNet, an online lexical database developed at Princeton University (Miller 1995). A word form (a string of letters) and a sense form a pair – more than 166,000 word form / sense pairs in total (Miller 1995: 40), but word forms may have many senses. The main sense of the word *ball* would be ‘a round object’, but other senses would be ‘globe’ or ‘lavish formal dance’. If two words can be substituted in a particular context, they are synonyms. Different senses will be linked to different synonyms. Sets of synonyms, so-called “synsets”, are the basic unit of WordNet (40). Under a synset, all words that can be used to express the same concept in the same context are subsumed, e.g., car, automobile, and vehicle. This way, synonyms are organized into groups; the semantic relations synonymy, antonymy, hyponymy, meronymy, troponymy, and entailment are encoded between the individual words. Open-class words (nouns, verbs, adjectives, adverbs) are distinguished, but closed-class words are not part of the data. Essentially, this approach results in a large semantic network. It constitutes a dictionary that is not structured alphabetically but according to meaning relations. A word is given along with an explanation as well as an example sentence. The database may therefore be used as a thesaurus, or even as an educational tool for language learning. Miller (1998: 45) reports that computational linguists showed greater interest in the project than psycholinguists. Whereas the latter see the approach as something that needs to be tested on the grounds of psychological plausibility, the former appreciate it for its usefulness and potential in the field of computer-based research. This explanation concurs with the interest in WordNet as a tool to imitate a semantic system in reading aloud.

Even though many semantic relations have been encoded in WordNet, the list is not exhaustive. Hyponymy is seen as the most important relation (Miller 1998: 24); a robin is a bird, a bird is an animal, and an animal is an animate being. A hyponym shares the features of its hypernym: It is not necessary to explain that a robin is an animate being; the links between these entries is enough (25). Miller (1995: 40) states that relations were chosen because "they apply broadly throughout English and because they are familiar – a user need not have advanced training in linguistics to understand them". However, this statement is not a fitting explanation, seeing that the most important relations that have not been considered are co-ordination and collocation. These two relations are by no means hard to grasp, but they may be hard to implement in a lexical database. One aspect of co-ordination, namely antonymy, has been realized, but other forms of word pairs or word lists (such as *salt* / *pepper* or *blue* / *green* / *red*) are left out. Collocations, which are an essential part of language, are likewise absent from WordNet. The database does not make any claim to be representative of the workings of the human brain, nor does it claim to be an all-encompassing approach. It is a useful tool that can be used to accomplish certain objectives.

After studying WordNet, it is worthwhile to take a look at how this database has been used in PDP. Harm (2002: 6-8) points out that there have been a number of different approaches to create semantic features, e.g., the experimenters created semantic features themselves, or study participants were asked to write down features for a list of words. However, such approaches are time-consuming, and participants will not list all possible features, but only those that they can think of and put into words (7-8). The database WordNet, by contrast, is fairly comprehensive, and it saves the researcher the time of constructing a resource from scratch. Another reason for choosing WordNet, and a semantic features approach in general, is that semantic features fit well into a model using distributed representations. Most features apply to many words (there is more than just one bird in the database), and features that belong to one word only were deliberately avoided (23). It should be noted that WordNet is only the basis for the semantic features that Harm & Seidenberg (2004) used in PDP, and that it needed some adaptation to arrive at a workable solution. Essentially, the relations hyponymy (a *canary* IS A *bird*) and meronymy (a *dog* HAS PART *tail*) were used to create the semantic features (Harm & Seidenberg 2004: 676). Only nouns and verbs received their features from WordNet, adjectives and adverbs were hand-coded instead (Harm 2002: 14). Harm (2002: 14) also states that verb features take a more simple form than nouns in WordNet; to take an example, *heat* and *improve* are both placed in the category 'change verbs', but there is no mention that heat refers to a (positive) change in temperature. To improve the list of features, pruning was carried out, and an additional set of features was added to make the meaning of words clearer and to disambiguate between words, e.g., [informal], [archaic], [negative], or [high-intensity] were added. Since WordNet does not contain any morphological features, features such as [plural] and [past tense] were added as well (Harm 2002: 15-17). The most profound change or simplification involves the plurality of senses encoded in WordNet: for use in PDP, only the most common sense of a word was kept (Harm 2002: 14), while all other senses were removed. To operate with such a wide array of senses was probably considered impracticable.

From the considerations presented in this sub-chapter, it is obvious that meaning evades any attempt at pinning it down. Other domains such as phonology can be more easily specified. The relation between the signifier and the signified is complex, and a signifier may have many senses. There are many ways to conceptualize meaning, and meaning is not only defined by what is said, but also by a word's relations to other words. The process of meaning-making is an exclusively mental activity, which makes it so difficult to adapt it for practical purposes. As Harm (2002) makes clear, WordNet is not a perfect copy of our psychological reality, but it is a workable and economic tool that enables a demonstration of how semantic relations could be activated in our mind while reading. The steps that were taken to adapt WordNet for integration into PDP have been illustrated; further details may be obtained by consulting Harm (2002).

5.3 The influence of semantics on reading aloud

After looking at meaning and its conceptualization, the next step is an investigation of how meaning relates to reading aloud. Is meaning just a byproduct of reading aloud, or does it influence reading, and if so, in what way? Overall, the limited interest in this connection demonstrates that the influence of semantics on the process of reading aloud is considered to be marginal. There have been very different ways of assessing a possible contribution of semantics. This sub-section will investigate three approaches: a study of patients suffering from semantic dementia (i.e., bleached conceptual knowledge (Patterson, Nestor & Rogers 2007), a study of the influence of a word's imageability on its naming time (Strain, Patterson & Seidenberg 1995) and a nonword training study with made-up words that were taught with or without meaning (McKay et al. 2008).

The first approach that will be examined here is the study of semantic dementia – a neurodegenerative disease – in relation to reading. Patients who suffer from semantic dementia have difficulty in remembering how an object or thing looks like. To better understand this, it should first be explained that scientists distinguish between two general types of memory: semantic memory and episodic memory. Patterson, Nestor & Rogers (2007: 976) define semantic memory as “the aspect of human memory that corresponds to general knowledge of objects, word meanings, facts and people, without connection to any particular time or place” and also refer to it as “conceptual knowledge”. Europeans know that a camel has a hump, that a duck has a beak, that a zebra has stripes, or that Monday is the first day of the week. Episodic memory, by contrast, refers to “specific events or episodes occurring in a given place at a given time” (Eysenck & Keane 2010: 255). This could refer to a chat with a friend in a café, or to visiting a football match. In semantic dementia, only semantic knowledge is affected, whereas episodic memory remains unaffected (Patterson, Nestor & Rogers 2007: 978). Patients forget the defining characteristics of a particular animal, but they still know what happened at last year’s birthday party. It seems that as a result, semantic dementia patients are unable to name the animal or object in question, even if a description is provided (978). In extreme cases, they no longer remember the broad category in which the object or animal belongs. Graham, Patterson & Hodges (2000) investigated the performance of semantic dementia patients on single word reading and spelling and found pronounced reading deficits, with irregular words being affected most strongly. In general, performance was influenced by regularity and word frequency (156), whereas performance on nonword reading was relatively normal (157). Graham, Patterson & Hodges (2000: 157-158) attribute the poor performance to the participants’ degraded semantic memory, as reading performance was strongly connected to the performance on semantic tasks. It appears that in semantic dementia, all information pertaining to a word is affected: orthography and phonology are on the same level as semantic features, as if all these aspects were petals of the same flower. It also seems that semantic dementia patients had less difficulty in reading aloud regular words and nonwords, which supports the existence of a rule-based GPC system.

Strain, Patterson & Seidenberg (1995) focused on imageability as a variable that possibly influences single word reading. The basic idea of this research project was that words might be read aloud faster if they are more easily conceivable in the mind. Words with low imageability, such as *cache*, were compared to words with high imageability, such as *comb*. Moreover, the variables regularity (regular vs. exception words) and word frequency (high vs. low) were included in the analysis. When looking at naming times of regular words, Strain, Patterson & Seidenberg (1995: 1143) found a (non-significant) difference of 5 ms between the groups. By contrast, the difference in naming latencies of high-frequency and low-frequency exception words was substantial, namely 40 ms (1143). Overall, words that readily evoke visual associations were read aloud faster than words that describe abstract concepts, but Strain, Patterson & Seidenberg (1995: 1144) observe that it is mainly low-frequency exception words that benefit from the interaction between regularity and imageability. The underlying argument is that the reading of these words takes longer and therefore allows for the effect to appear. Strain, Patterson & Seidenberg (1995: 1147) conclude "that representations of word meaning are activated in the course of orth-to-phon translation and […] that semantic representations for imageable words make a useful contribution to this computation in the case of low-frequency exception words". This view fits well with the petal metaphor mentioned earlier: a larger number of petals is active for highly imageable words, which leads to faster naming. However, it should be stated that there was no reliable three-way interaction between naming time, word frequency, and imageability (1144), and that imageability is a subjective concept that is hard to measure.

A somewhat creative approach to researching the influence of semantics on reading was taken by McKay et al. (2008), who set out to investigate the role of semantics by means of a nonword training study. Fictional words were taught to a group of participants; in the "nonsemantic condition", only the orthographic and phonological forms of these made-up words were taught, in the "semantic condition", a meaning was provided for each word as well (McKay et al. 2008: 1497). The semantic condition included a sentence-completion task in the training stage. A greater number of inconsistent stimuli was included on the assumption that the error rates for these words would be higher; moreover, inconsistent words were inconsistent in the sense that their vowel was basically plausible, but in fact unattested in similar orthographic bodies (1497). It was found that naming in the semantic condition was faster and more prone to be error-free, but only for nonwords with inconsistent pronunciations (1508). However, this effect was only observed in the second experiment in which more emphasis was placed on simulating real-life lexical acquisition: the words' phonology and their meaning were taught prior to their spelling (1504). Again, it appears that semantics, phonology, and orthography form a complex whole, and that semantics has at least some influence on naming latencies.

Considering all evidence presented, it may be concluded that the influence of semantics on naming latencies tends to manifest itself when reading aloud exception words. Regular words, which may be read based on a rule system alone, do not seem to be affected by the activation of meaning. Possibly, regular words are read aloud so fast that semantic information only reaches full activation after the word has been pronounced, or is in the process of being pronounced. For exception words, naming takes longer, allowing semantics to ‘kick in’ and boost naming speed. This would lead to faster naming of imageable words, which have a richer semantic representation than their low-imageability counterparts. Evidence from semantic dementia patients suggests that phonology, orthography, and semantics are closely connected: if one aspect of a word’s information is impaired, other aspects are likely to be impaired as well. This interesting realization was illustrated with the petal metaphor.

6. Experimental part

The experimental part of this thesis will link the insights obtained from the literature review to real-life participant data. The central aim is to investigate the role of meaning in reading aloud. Reading aloud has a clear input, i.e., printed letters, and a defined output, i.e., sounds, even though the exact realization (intonation, prosody) may differ. The large majority of studies only investigate the relationship between these two dimensions: orthography and phonology. While everyone will agree that the process of meaning-making is important, meaning is rarely addressed in scientific studies due to its elusive nature. Indeed, it is hard to say what constitutes the process of meaning-making. I should therefore emphasize that the nature of this project is largely exploratory. There are no ‘die-hard’ answers to the questions posed in this paper (even though some sub-questions may have definite answers). Still, the observations made have led to interesting realizations that have allowed me to form a picture of what goes on in our minds while we read.

6.1 Research design, research questions and hypotheses

From the literature review, it is clear that PHs have the potential to activate aspects of meaning, either through their phonology (which they share with real words), or through orthographic similarity to their base-words or similarly spelled words. Yet, the aspect of meaning has rarely been addressed by earlier work in this area. This paper aims to investigate the following overall question:

To what extent is meaning activated when reading aloud PHs?

Perhaps the activation of meaning in PH reading has been taken for granted, or else the researchers did not expect to gain anything from this investigation, or perhaps they just did not know how to tackle the question. Indeed, setting up a research design for this purpose is complicated. For one thing, the list of potential stimuli whose meaning is straightforward is short. Many words have more than one meaning, a dominant meaning and one or more other meanings, and these meanings may be related or unrelated. And since it is hard to pinpoint meaning, there is no obvious methodology to investigate comprehension in single word reading.

In coming up with a research design for this project, the use of pictures in combination with words was a central idea. It is commonly known that a picture says more than a thousand words, and therefore, it seemed intuitive to confirm the understanding of words through the presentation of images. While pictures do not represent meaning as such, they act as strong signifiers (in Saussure's terminology) that trigger the same signifieds as word stimuli. Both letters and image may activate similar aspects of a word's meaning, such as sensory experience (e.g., taste or smell) or personal associations. Thus, a participant's response to a picture may indicate if the signified was triggered when the letter combination was presented. Finding suitable pictures for such an undertaking is a challenge, as there are likely no images that the majority of people would identify as truly prototypical.

In Section 4 of this paper, the literature review has shown that PHs were usually compared to nonwords, but rarely to real words. Grainger, Spinelli & Ferrand (2000) compared real words and PHs in a pure block design, but for the investigation of mixed lists, they switched to nonwords instead of real words. It would therefore be interesting to see how PHs and real words perform in a mixed list design with regard to a word frequency effect. Such a design would complement the existing literature; moreover, the comparison of these two word types lends itself to the investigation of the research question of this paper, seeing that they both have the potential to trigger meaning, whereas nonwords are unlikely to do so. Studies employing a pure block design found a frequency effect for PHs (Grainger, Spinelli & Ferrand 2000; Borowsky, Owen & Masson 2002; Reynolds & Besner 2005), but in a mixed list design, the effect was rarely attested. The overall research question of this paper will be complemented by 6 more specific questions; one of these will investigate if a frequency effect occurs in a mixed-list design based on PHs and real words.

Seeing that a word frequency effect should be investigated, it made more sense to conduct a separate task for word naming. The presentation of pictures in between the words would constitute a further distraction, resulting in a 'double-mixed' design that is likely to change the outcome of the experiment (mix of two word types, plus a mix of words and pictures). In the experiment, participants will first perform a standard naming task. To get a better idea if they have understood what they were reading, it was decided that after the end of the task, the participants would be asked which words they remembered. This way, they would not focus on meaning-making during naming and would give an unbiased pronunciation, as they would not know that meaning was of importance. The second task will then combine naming and pictures; participants are asked to indicate if the picture matches the word that they have just read (I will call this a forced-choice picture task, as the answer is either yes or no). Correctness of response and reaction time to the pictures will be recorded. While the first task focuses on naming and pronunciation, the second task focuses on word meaning. In addition, eliciting participants' thoughts after taking part in the experiment is supposed to shed some light on PH processing.

I will now go on to outline the methodology. The experiment was conducted in individual (one-on-one) sessions in a quiet room on a netbook type HP 2140 Mini-Note with full-size headphones connected to it. Prior to performing the two tasks, participants were asked a few background questions: age, gender, mother tongue, other languages spoken, presence of normal vision, and known reading disorders. The target group was native speakers of English who had normal or corrected to normal vision and did not have any known reading disorders. Considering the sample size used in similar experiments, 20-25 participants was defined as an appropriate sample. After answering the background questions, a practice trial with five stimuli was performed to give participants a certain level of confidence. For the practice trial as well as the main experiment, the stimulus presentation software DMDX Version 5 (Forster & Forster 2003) was used to show letter combinations and pictures on the netbook's screen. The use of stimulus presentation software is commonplace in psychological experiments; it allows the researcher to present stimuli in a randomized manner and automatically records the participants' response with millisecond accuracy. This response can be verbal, or else the participant presses a button on the keyboard or on a button box. More information on DMDX will be given in sub-section 6.1.4.

6.1.1. Naming task

The first task – a naming task – required participants to read aloud a list of single word stimuli that appeared on a computer screen in a large font size (black letters on white background). Real words were randomly mixed with PHs; each stimulus appeared on the screen for two seconds. Participants were instructed to read the words aloud swiftly once they appeared on the screen. Verbal responses were recorded as single word files (.wav) for later analysis. After the participants had finished reading, they were asked which words they remembered. This question had not been announced prior to the start of the experiment, as prior notice might have influenced the outcome. All words that were named by a participant were written down on a sheet of paper by the experimenter.

The naming task consisted in reading aloud single words presented in isolation. For each group of words (real words and PHs), 25 words were selected to serve as stimuli, resulting in a total of 50 words. The two word lists were equated on number of letters, (base-)word frequency per million according to the $\mathrm{SUBTLEX}_{\mathrm{US}}$ corpus (Brysbaert & New 2009), number of orthographic neighbors (CLEARPOND Database, Marian et al. 2012), initial phoneme and wherever possible second phoneme (categories: affricates – fricatives – plosives – nasals – vowels – glides – liquids) and imageability (MRC Psycholinguistic Database, Version 2, c.f., Wilson 1988, based on Version 1 by Coltheart 1981). On the whole, it turned out to be a challenging venture to find enough highly imageable words to create two matching lists – this is clearly the limiting factor in this experiment. Due to this limitation, it was only possible to equate the summed characteristics of a group; to match every word of Group A with a word in Group B turned out to be an unrealistic goal in this research design. The PH stimuli were taken from prior studies involving PHs (Martin 1982; Taft & Russell 1992; Marmurek & Kwantes 1996; Herdman et al. 1996; Seidenberg et al. 1996; Borowsky, Owen & Masson 2002; Reynolds & Besner 2005), as these stimuli had been tried and found suitable for this kind of experiment. The real word stimuli were chosen with the aim to match the PH word list as stated previously. The final list comprises only nouns, as nouns are most easily representable by a picture (cf., picture task).

Some stimuli had to be excluded from the final list for reasons pertaining to the area of semantics. Due to the fact that native speakers currently residing in Vienna acted as participants in this study, interlingual PHs had to be dropped (e.g., *bote* would be a PH of *boat*, but means *messenger* in German; *gift* was intended as a real word match for *caik*, but *gift* translates into *poison* in German). The word *tode* could be taken to mean *deaths* in German, but the plural form is extremely rare and could be considered unusual. Words that had the same meaning were kept, as a German pronunciation would not have led to a different response in the picture task (e.g., *ball*). A few words were excluded if there was a homonym with a clearly unrelated sense (*wale* could be interpreted as *whale* or *wail*). A more lenient position was adopted in case of related senses (e.g., *orange* could stand for both the fruit and the color). Another factor to be considered were plural words, as they add an extra dimension to the understanding of words in reading. Therefore, only words in their singular form were chosen.

6.1.2. Forced-choice picture task

The forced-choice picture task combined words (naming) with pictures. Naming was identical to the setup described above, but this time, each word was immediately followed by a picture. Like the letter combinations, the pictures were presented for a duration of 2 seconds. After the participant had read aloud the letter combination, a picture appeared, and the task was to indicate if the picture matched the letter combination or not by pressing one of two buttons. For clarification, it was mentioned that the letters *gerl* followed by the picture of a girl was considered to be a match. A word and a picture formed a unit; these units were presented by the software in random order. All pictures used were colored clipart; the decision in favor of colored rather than black-and-white images was meant to make objects appear natural. Moreover, the color of an object is often a defining feature, and thus contributes to the activation of meaning. In general, decently colored clipart was chosen from open sources. The aim was to select natural, true-to-life looking clipart; the use of cartoonish clipart was avoided (pictures are given in the Appendix). Stimuli were matched for resolution and saved as bitmap files. Picture size was harmonized as well (15 cm width, 17 cm in case of pictures with low height and 10 cm in case of pictures with large height). Each picture was paired with a word or PH; there were three possibilities:

- a letter/picture combination was a match
  e.g., *berd* was followed by the picture of a bird
- a letter/picture combination was semantically related
  e.g., *aks* was followed by the picture of a hammer
- a letter/picture combination was semantically unrelated
  e.g., *bench* was followed by the picture of a goat

In each group (words and PHs), 13 letter/picture combinations were a match and 12 combinations were a mismatch. Out of these mismatches, 6 pairs were semantically related and 6 pairs were semantically unrelated. Not every word lent itself equally well to replacement by a semantically related object. As a first step, a list of possible semantic replacements was created, 6 word pairs were then chosen from this list. The replacement of 6 word pairs by semantically unrelated objects was easier insofar as random objects could be used as a replacement. In choosing the pairs, care was taken to match the groups on initial phoneme, e.g., *beer* and *boan* were both part of the semantically unrelated group. Moreover, the base-words of the semantically related or unrelated pictures had a similar summed word frequency per million, e.g., 238.09 (words) versus 237.28 (PHs) in the semantically unrelated condition.

The following word / picture pairs were chosen:

- semantically related
  soope / pizza, *hownd* / wolf, *phish* / (sea)turtle, *aks* / hammer, *angker* / ship, *sord* / shield, *skull* / grave, *hedge* / tree, *fruit* / carrot, *egg* / chicken, *owl* / eagle, *orange* / apple
- semantically unrelated
  fone / king, *spunj* / bomb, *boan* / snake, *sircle* / bag, *rane* / cake, *brume* / scarf
  river / bride, *bench* / goat, *face* / corn, *beer* / fan, *spear* / pen, *sunset* / king

As this list shows, the semantic relation between word stimulus and picture was largely co-hyponymy (*soup* and *pizza* belong to the category food, *fish* and *sea turtle* are water animals, *axe* and *hammer* are tools, *owl* and *eagle* are birds, *orange* and *apple* are kinds of fruit, *sword* and *shield* are both used in battle, *hound* and *wolf* are canine, and *hedge* and *tree* are plants). As such, they come from the same semantic field. Other relations were meronymy (*anchor* / *ship*) and connotation (*egg* / *chicken*, *skull* / *grave*).

6.1.3. Research questions and hypotheses

Now that the tasks have been explained, I will present the more specific research questions that will be analyzed as part of this research, along with the corresponding hypotheses.

Naming task:

a. *Is there a difference in naming latencies between real words and PHs?*
   In accordance with the literature, real words are read aloud faster than PHs.
b. *Is there a (base-word) frequency effect for real words and PHs?*
   In accordance with the literature on mixed list experiments, it is likely that there will be a base-word frequency effect for real words, but not for PHs.
c. *After reading aloud, do participants remember more real words or PHs?*
   Participants will remember more real words than PHs because of their familiar orthographic representation. PHs might only be seen as an obstacle to reading.

Picture task:

d. *Is there a difference in response time between real words and PHs?*
   Due to the unfamiliar nature of PHs, responses to real words will be faster than responses to PHs.

e. *Is there a difference in the number of correctly identified pictures between real words and PHs?*

   Due to the unfamiliar nature of PHs, there will be a larger number of correctly identified pictures following real words.

f. *Is there a difference in response time between the conditions 'semantically related' and 'semantically unrelated', both for real words and PHs?*

   Responses to semantically related pictures will take longer than responses to semantically unrelated pictures (because they create a bigger conflict due to overlap of semantic features).

The first two questions are very common in similar experiments, and could therefore be categorized as basic questions. The dependent variable is naming time, which is put in relation to the independent variables word type and word frequency. The third question is an open question that has not previously been employed in this kind of experiment. The question probes if the participant has indeed understood the PH stimuli while reading them, and not just pronounced the letter combination swiftly in order to move on to the next word.

The picture task is a new research design altogether. The main variables are correctness of response and response time. Both variables are dependent variables in the sense that they depend on semantic activation. Only if the word's meaning has been activated can the participant give a correct response (or else, s/he would have to rely on chance), and the time taken to respond will either be slower or faster, depending on the recognition of the signified. The last question investigates the difference in response time between semantically related and semantically unrelated pairs. It seemed interesting to assess this difference, as a longer response time for semantically related words might indicate the presence of a greater conflict.

6.1.4. Stimulus presentation and analysis software

These days, researchers may choose from a variety of stimulus presentation software. Some of these programs are freely available, others require payment of a license fee. Software may vary in complexity; more complex research designs may require a large set of options, but to fully make use of these options may necessitate programming skills. DMDX display software (Forster & Forster 2003) was chosen for three main reasons: it is freely available, it is fairly easy to set up (and does not require knowledge of programming languages in doing so), and it has been used in naming experiments. Essentially, it measures reaction times to visual and auditory stimuli with millisecond accuracy (Forster & Forster 2003: 116). The experimenter needs to prepare an item file or script that contains a header line and a list of items. In the header line, a number of parameters is specified, such as method of response (keyboard, response box, mouse, headphones, voice key), standard frame duration (how long a stimulus is displayed), delay from the end of one item until the next one is presented, time-out (how long the program waits for a response by the participant), information on item scrambling, background color of the screen, font color, and font size. Depending on the research design, the number of required parameters may vary. The list of items tells DMDX what to present, e.g., text, pictures, video, or sound. Text stimuli are entered directly into the script in brackets. Typically, the script begins with instructions and ends with a thank you note. In the main text, every item receives an item number, e.g., 101 for the first item in the first group of words. A plus sign in front of the item number means that 'yes' is the correct answer, a minus sign means that 'no' is the correct response, and an equals sign means that any response is correct. In a naming task, there are no correct responses, only a verbal answer that is recorded. An example of a text item in a script is =107 * "noze". For familiarization with DMDX, an online tutorial by Isabelle Darcy (2010) based on an earlier version by Matt Davis (1999) is available.

Experiments involving word naming typically make use of a device called a voice key. The purpose of a voice key is to register the onset of articulation, and to communicate this time (in ms) to a software that logs it for later analysis. The device is connected to a microphone; it contains a switch that is activated in accordance with a pre-set acoustic activation threshold. Once the sound level rises above this threshold, the device logs the time that has elapsed since the stimulus appeared on the screen. As observed by Kessler & Treiman (2002), voice key measurements need to be treated with caution. The main factor complicating the logging of response times is the number of false or inaccurate starts due to non-linguistic noise (e.g., coughing) or participants speaking in a low voice. There are various ways to remedy this issue; the most efficient one probably consists in avoiding the use of a voice key by recording the responses as audio files and extracting the reaction times with the help of software. In case of doubts, the experimenter may visually inspect the waveform individually in the respective audio file.

Instead of a voice key, the software CheckVocal (Prototapas 2007) was used to analyze the naming task. Other options were considered as well; the first option was SayWhen (Jansen & Watter 2008), but unfortunately, this software does not work with single sound files. Another option was an online resource named Chronset (Roux, Armstrong & Carreiras 2017), but this tool was found to be unreliable after comparison with a 'manual' analysis of the file.

CheckVocal has the advantage that it is automated, but allows for manual correction. As a first step, the experimenter needs to create a file that includes the correct responses; the software then shows this response text and provides both an oscillogram and a spectrogram of the participant response. CheckVocal can be instructed to show a red line at the presumed voice onset; this line can then be moved by the experimenter in accordance with the spectrogram and oscillogram. It is possible to play the sound to the right and to the left of the red line, which helps the decision process. Figure 8 shows an example of the analysis of a word via CheckVocal.

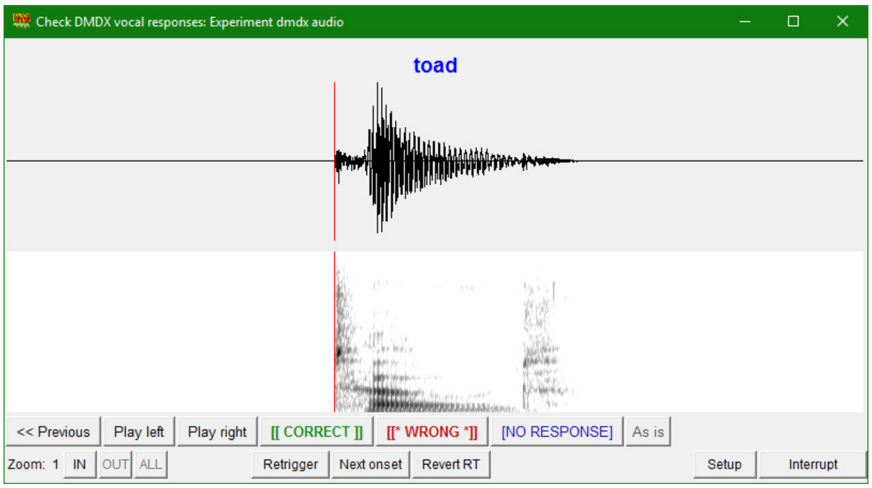


Figure 8 Analysis of the naming task via CheckVocal

In Figure 8, the onset of the word is triggered correctly; the red line does not have to be moved. In addition to determining word onset, every response has to be marked as correct or wrong. If a word was pronounced as intended, the experimenter marks the response as correct; otherwise, the response is marked as wrong. All stimuli are presented one by one, but the process is still reasonably fast and the result is very likely to be accurate. Some initial phonemes are more tricky to analyze than others, e.g., in the oscillogram it is fairly difficult to say when the sound /f/ starts (/f/ is mistriggered most frequently, followed by /h/ and /s/), and the onset of /b/ can show great variation in length. At the end of the analysis, the software creates a new file that gives the participant ID, date and time, and all naming latencies in one row. Wrong responses are marked with a minus in front of the response time. Responses are not presented in their random (scrambled) order as they were shown to the participant, but are listed in the original order (101, 102, 103, etc.), which is very helpful.

The forced-choice picture task was analyzed with the tool azk2txt which is supplied along with CheckVocal. This tool does not necessitate any experimenter input; it creates a file for each participant that lists participant ID, date and time, and all response times in one row. Wrong responses (i.e., the participant indicates that the picture matches the word, even though this is not the case) are preceded by a minus. The main purpose of using azk2txt is to get the responses in a neat form, presented in the original order. Both CheckVocal and azk2txt only work if a participant ID has been keyed in for every trial. If a participant ID is missing, automatic analysis cannot be performed, and a manual analysis of the data can be quite tedious.

Every software inevitably requires a certain amount of familiarization on the part of the researcher. Clearly, a script is unlikely to work well from the outset, but will need reworking and fine-tuning. Millisecond measurements would not be possible without the use of computer software, and while the analysis of a task can be performed without specialized programs, it would take much longer. Initially, finding and using suitable software is time-consuming, but it is well worth the effort. In short, stimulus presentation software is both a blessing and a curse.

6.1.5. Factors to be controlled for in pseudohomophone experiments

In naming and lexical decision experiments, two or more groups of words are presented to the participants. It is understood that these groups of words should be roughly equal – there should be no significant differences with regard to a number of features, such as the number of letters of a word. This process is called ‘controlling for a number of factors’. The factors that have been proposed as requiring such a control procedure are manifold (letter length, word frequency, number of orthographic neighbors, bigram frequency, to name but a few); to control for every single one of these factors would be close to impossible. In practice, researchers should therefore look at the type of their experiment and determine which factors are the most important ones that are likely to influence the outcome of the study, and make their provisions accordingly.

The fact that letter length and word frequency have been shown to influence word naming has been mentioned earlier in this paper. The theory is that a word takes longer to name the more letters it has, and that highly frequent words are named faster than words of lower frequency. While word length is easy to determine, word frequency has to be looked up in a database. For the present experiment, word frequency was determined with the help of the $\text{SUBTLEX}_{\text{US}}$ corpus (Brysbaert & New 2009). This corpus is based on word frequency as it occurs in the subtitles of US-American television series and films and is freely available online. Another factor that has already been explained is the number of orthographic neighbors of a word (sometimes called Coltheart’s N). If there are many words that are similarly spelled as a particular word, this word might be pronounced faster because it receives greater activation. As it would be difficult to determine the number of orthographic neighbors solely with the help of a dictionary, the CLEARPOND database (Marian et al. 2012) was used. This resource makes it possible to look up the neighborhood characteristics of words in 5 different languages. Factors that are related to neighborhood density are bigrams (also called digraphs) and trigrams. This means that certain combinations of two or three letters are relatively more frequent than others. Even though it might seem thorough to control for these frequencies as well, it was decided that this would go a bit far and would also be hard to accomplish due to the limited number of suitable word stimuli. Therefore, bigram and trigram frequency have not been controlled for in this experiment, and the number of bigram or trigram neighbors has not been considered either.

A factor that surely ought to be taken into consideration in a study of reading aloud is the physical realization of sounds. Sounds are likely to be pronounced more or less swiftly depending on the manner of articulation of their initial phonemes, simply because the movements of the articulators take longer to perform. According to Kessler & Treiman (2002: 159), response time varies in the following way (from slowest to fastest): affricates – fricatives – plosives – nasals – vowels – glides – liquids. This order is hardly surprising, seeing that more effort is required to produce an affricate than a liquid. As a result, the groups of words used in a naming experiment need to be matched for initial phoneme type according to these seven categories, i.e., to include an equal number of words that begin with, e.g., an affricate or a vowel.

The last (but not least important) factor that was determined in the preparation of this study is imageability. Naturally, only words that are sufficiently imageable were suitable candidates for use in the experiment. It was also mentioned that imageability is a somewhat subjective construct. Still, it is possible to determine the imageability of words by asking participants to rate a word on a scale. As it would be time-consuming to conduct such a survey, existing data from the MRC Psycholinguistic Database (Version 2, c.f., Wilson 1988, based on Version 1 by Coltheart 1981) was used to corroborate the high imageability of a word. In this database, it is possible to select characteristics (e.g., a four-letter word of a certain frequency) and receive a list of highly imageable words. The imageability rating ranges from 100-700; all words used in this study have a rating of 494 or above.

It has also been proposed that the semantic properties of a word affect the speed at which the word is accessed, the basic idea being that words which activate more semantic information are accessed faster. Pexman (2012) provides a concise summary of what she refers to as "semantic richness effects" (Pexman 2012: 31) that apply to visual word recognition. Such effects may be caused by:

- imageability
- number of features (e.g., 'has a tail')
- number of words that are associated with a word (e.g., *lace* → dress, pretty, frills, …)
- semantic neighborhood (number of collocates; this can be determined by using a corpus)
- body-object interaction (some words may invoke memories of sensorimotor experience, such as touching or smelling the object, but also emotional reactions)

Other influencing factors would be ambiguity and valence (the kind of emotion that a word triggers, positive or negative associations).

Seeing that a possible influence of semantic properties on word naming has to date not been considered in naming experiments, all factors except for imageability have been disregarded in the present experiment as well. However, Pexman (2012) has a valid point in highlighting the semantic dimension in the recognition of words.

Table 3 Word stimuli used in the experiment

			letters	frequency p.m.	initial phoneme	Coltheart's N	imageability
1	sircle	PH	6	21.51	fricative	1	591
2	berd	PH	4	45.45	plosive	9	614
3	rane	PH	5	48.90	liquid	24	618
4	brane	PH	5	77.02	plosive	8	572
5	boan	PH	4	26.06	plosive	12	567
6	chare	PH	5	49.24	affricate	10	610
7	noze	PH	4	69.75	nasal	8	605
8	soope	PH	5	25.20	fricative	2	604
9	focks	PH	5	21.61	fricative	12	607
10	fone	PH	4	269.73	fricative	15	587
11	hownd	PH	5	5.04	fricative	1	596
12	phish	PH	5	83.49	fricative	1	615
13	bair	PH	4	57.41	plosive	9	572
14	tode	PH	4	5.69	plosive	14	591
15	kee	PH	3	86.86	plosive	19	618
16	angker	PH	6	7.41	vowel	2	561
17	aks	PH	3	4.88	vowel	7	597
18	kanoo	PH	5	3.57	plosive	1	602
19	gittar	PH	6	15.59	plosive	0	unavailable
20	ere	PH	3	32.00	vowel	13	597

21	spunj	PH	5	6.71	fricative	2	577
22	sord	PH	4	26.18	fricative	9	597
23	leef	PH	4	5.20	liquid	5	608
24	baik	PH	4	25.88	plosive	6	unavailable
25	brume	PH	5	4.76	plosive	2	608
				1,025.14		192	596.26 (mean)
26	sunset	RW	6	10.31	fricative	1	633
27	ball	RW	4	104.96	plosive	19	622
28	river	RW	5	55.47	liquid	9	633
29	bread	RW	5	28.33	plosive	9	619
30	beer	RW	4	75.49	plosive	13	598
31	chain	RW	5	21.22	affricate	8	559
32	mask	RW	4	19.80	nasal	11	unavailable
33	skull	RW	5	14.71	fricative	2	609
34	fence	RW	5	16.06	fricative	3	611
35	face	RW	4	289.16	fricative	15	581
36	hedge	RW	5	1.55	fricative	5	583
37	fruit	RW	5	21.73	fricative	2	587
38	book	RW	4	176.98	plosive	16	591
39	tire	RW	4	12.37	plosive	14	511
40	cup	RW	3	51.65	plosive	14	558
41	orange	RW	6	22.31	vowel	3	626
42	egg	RW	3	26.04	vowel	2	599
43	camel	RW	5	5.02	plosive	4	561
44	garlic	RW	6	6.00	plosive	1	565
45	owl	RW	3	5.61	vowel	7	595
46	spear	RW	5	4.55	fricative	7	545
47	sofa	RW	4	5.86	fricative	3	597
48	lamb	RW	4	10.63	liquid	7	614
49	belt	RW	4	24.35	plosive	10	494
50	bench	RW	5	9.67	plosive	5	555
				1,019.83		190	585.25 (mean)

Frequency p.m. – SUBTLEX$_{US}$ database; number of orthographic neighbors or Coltheart's N – CLEARPOND database; imageability – MRC psycholinguistic database; RW = real word

Table 3 lists the stimuli that were used in the experiment along with their characteristics (word type, number of letters, word frequency, initial phoneme, number of orthographic neighbors, and imageability). In addition, the total of each group is given at the bottom.

6.2 Piloting phase

The piloting phase provided valuable insights. The assumption that a practice trial would be useful in giving the participants practice and confidence prior to performing the actual task was definitely confirmed. The anxiety that was present at the very beginning of a session soon dissipated. Moreover, piloting gives the experimenter a certain level of routine in administrating the task to the participants (providing clear and sufficient information, remembering to type in the participant ID that greatly facilitates analysis via CheckVocal and azk2text). Primarily, piloting was meant to show if the presentation time of a stimulus was sufficient, too short, or too long, and if the word stimuli were pronounced as intended, and if the picture stimuli were perceived as fitting.

Initially, the experiment was piloted with students of English to iron out the main bugs before presenting it to native speakers, whose availability was more limited. Presentation time turned out to be sufficient if not a bit too long, and the pictures were perceived to be suitable representations. The fox was perceived to be rather cute, which caused a participant to smile. A hound was perceived to be very similar to a wolf (not surprising in the semantically related condition), and one person reported that she did not really have an image in her mind for a hound. Another surprising comment was that the picture of a snake had created confusion for the word *sircle*. The participant commented that the winding form of the snake had caused this confusion, which must have reminded her of the circle, but upon pondering it further, it seemed more likely that *sircle* must have activated the representation of *serpent* – a case of phonological distraction because of the identical onset. To avoid this type of confusion, the pair *sircle* / snake was broken up, and the picture of a snake was paired with a different stimulus.

Overall, testing was smooth and without technical complications. Even though English was not the mother tongue of the students that were tested, there were only minor mispronunciations of the PH stimuli (~5 out of 25, e.g., /'ɪəʳɪ/ instead of /ɪəʳ/). There were also minor 'L2 mistakes' among real words: *arch* was pronounced /ʌrk/ and *worm* was pronounced /wɔ:ʳm/ (these items belonged to the group of practice stimuli). Performance on the picture task was very good (0-5 mistakes). There were 2 comments on the pictures: that the camel might in fact be a dromedary (which was not the case) and that hound and wolf might actually be the same concept. One participant suggested to emphasize that the experiment was not a test; another participant suggested to mention that the interval between words was always the same, and that the next stimulus did not appear earlier if the answer was given faster.

7. Results

The data obtained by means of stimulus presentation and analysis software was stored and organized in Microsoft Excel; regression analyses were performed in R Studio (R Core Team 2015). Before presenting the answers to my research questions as listed in Section 6.1.3, I will present an overview of the figures that I obtained as part of my research. As will be explained in the sub-chapters, some figures look different if problematic items or long reaction times are deducted.

Table 4 Overview of main results

real words			PHs		
mean naming time per item in ms (N=20)					
555.51			644.61		
summed remembered items of all participants (N=20)					
46			106		
summed errors of all participants (N=19) in the picture task					
11			22		
match	related	unrelated	match	related	unrelated
2	7	2	6	14	2
mean reaction time to picture in ms					
710.46			737.67 725.03*		
for related pairs					
825.96			894.37 867.14*		
for unrelated pairs					
714.47			736.72		

* after removal of the item *hownd*

For the picture task, the summed errors of all participants are given. The maximum number of possible errors is 475 (25 word-picture combinations in each group multiplied by 19 participants). These errors are then divided between the three conditions (match, related, unrelated).

To give a better insight into the research process and to paint a more vivid picture, I will also give the findings of an individual participant. MH, a 25-year-old native speaker of American

English, had a mean naming time of 552.72 msec for real words and 638.88 msec for PHs. She took longer than one second to pronounce three items: *berd*, *bair*, and *gittar*. She remembered 6 PHs and 3 real words, namely *baik*, *bair*, *ere*, *kanoo*, *sord*, *tode*, *bread*, *garlic*, and *skull*. In the picture task, she only had one error for the pair *hownd* / wolf, and since she felt that a hound and a wolf were very similar, this was actually not much of an error. Surprisingly, she responded faster to PH / picture pairs (543.65 msec) than to real word / picture pairs (580.22 msec). This was the case with 7 out of 19 participants, albeit the difference was sometimes marginal. The pairs *aks* / hammer and *egg* / chicken showed the longest reaction times (both slightly above one second). Responses to semantically related pairs took longer (738.38 msec PHs / 708.18 msec real words) than responses to semantically unrelated pairs (530.47 msec PHs / 549.33 real words).

Having taken a look at the results of one specific person, it becomes clearer that the main difference between the two stimuli – real words and PHs – presents itself in reading aloud. The difference in the picture task is much less pronounced, and biggest for semantically related pairs. The most plausible explanation appears to be that the pronunciation of an unfamiliar word inevitably requires a greater effort, but once its sound pattern has been computed, it swiftly activates the base-word as well as other words that are related in any way. To show if this reasoning is accurate, a more detailed analysis will be presented on the following pages.

7.1 Description of sample

Finding participants who would take the time to participate in a one on one experiment was not an easy task. Some of the participants were recruited at university (students as well as teachers), others were recruited in internet forums for international residents in Vienna. It appears that most native speakers use their language skills also professionally; they work as (language) teachers, translators, editors, or copy writers. The mean age of the participants was 37.1 years; their age ranged from 21 to 64 years. In total, 14 women and 6 men took part in the study. Women were more likely to volunteer online, or to respond when they received a message by a stranger. 13 participants spoke American English, 2 spoke Canadian English, 4 spoke British English, and one person had a mixed accent. Their knowledge of German differed greatly, from little or no knowledge (in the case of a business traveler) to near-native German (three of them either had a degree in German or were advanced in their studies of German). Other languages spoken were often Spanish and/or French, but participants also spoke (some) Italian, Portuguese, Russian, Latvian, Hungarian, Nepalese, or Japanese. Three participants had grown up bilingually (2 English/German, 1 English/Finnish). All participants had normal or corrected to normal vision, and they did not have any known reading difficulties such as dyslexia.

Most participants carried out the tasks in a professional manner. The data of one person had to be dropped altogether because of extremely slow naming / reaction times which were possibly due to medication. The data of this participant has been excluded from the above description of the sample group. For the picture task, there was no data from one participant. The analysis of the task showed mostly timeouts, possibly due to very slow reaction times (during the experiment, everything appeared to be fine, apart from the fact that the person was left-handed, and the choice of yes/no buttons favored right-handedness). This left 20 participants for the analysis of the naming task and 19 participants for the analysis of the picture task.

7.2 Naming task

In the following, the three research questions pertaining to the naming task will be analyzed. These questions investigate the difference in naming latencies between real words and PHs, the presence or absence of a frequency effect for both word types, and the words that were most frequently remembered by the participants after reading.

a. *Is there a difference in naming latencies between real words and PHs?*

On the whole, most items were read as intended, but some turned out to be more problematic than others. It soon became evident that the item *baik* was most often pronounced 'bake' by native speakers of English, but it was also pronounced 'bike' by others. I had wrongly assumed that the item was meant to represent a PH of 'bike', and it appeared that participants with excellent knowledge of German had a tendency to interpret it the same way. Another problematic item was *ere*, which was not always pronounced 'ear', but also 'ere', as in the archaic word meaning 'before'.

Unsurprisingly, real words were named faster than PHs. After deducting all answers that differed from the intended pronunciation, and leaving the problematic items *ere* and *baik* out of the analysis, the mean naming time (N=20) for a PH amounted to 644.61 msec, as opposed to 555.51 for real words. This equals a mean difference of 89.1 msec per item. After also deducting all responses that were untypically slow (> 1 sec), this difference was even less pronounced (74.46 msec, 623.86 vs. 549.4 msec). Slow answers were given almost exclusively in response to PHs. For real words, there were only five slow readings in total, and only one incorrect reading (*beer* was pronounced as /bɛ^{r}/).

In the group of PHs, the items *aks* and *angker* took longest to pronounce. In all likelihood, this finding is due to an ensuing conflict, as these words are very similar to the real words *ask* and *anger*. In fact, *aks* was sometimes pronounced 'ask', and thus it seems that participants were torn between two conflicting pronunciations, which led to longer naming times. The same is true for *angker* and 'anger', even though with lower frequency. The words that were pronounced fastest were *leef* and *sord*, which is interesting because their base-word frequency is low (5.2 p.m. for *leef*) to mid-range (26.18 for *sord*). It would appear that in this case, visual similarity to the base-word exerted a positive influence.

Visual similarity has also been observed to play a role in nonword reading. In mixed lists, nonwords should theoretically be pronounced slower than PHs, e.g., *chief* 437 msec, *cheef* 545 msec, *bleef* 591 msec (Marmurek & Kwantes 1996: 23). However, nonwords are sometimes pronounced faster, and this is likely the case if a nonword's orthography is visually similar to that of a real word, e.g., *fox* 450 msec, *phocks* 808 msec, *snocks* 564 msec (> *socks*); *green* 432 msec, *grean* 601 msec, *drean* 567 msec (> *dean*, *dream*) (Marmurek & Kwantes 1996: 22). This shows that visual similarity can have a facilitatory effect, but it is only one of many factors that are at play in word reading. Other factors such as word length and initial phoneme have an influence as well, and the naming time of letters is therefore a result of the interplay of numerous factors.

As already stated, knowledge of German appeared to influence naming. Participants without or hardly any knowledge of German pronounced the items *baik* and *soope* as intended ('bake' and 'soup'), but participants with excellent knowledge of German (who were either students of German or had lived in Austria for a long time) tended to pronounce them 'bike' and 'soap'. It appears that our internal system for assembling unfamiliar words may be influenced by our knowledge of other languages, but only after considerable exposure to the language.

b. *Is there a (base-word) frequency effect for real words and PHs?*

To find a possible effect of word frequency on naming, a regression analysis was performed for these two variables. However, the results indicated that this approach might not be the best way to go about this task, as the range of frequencies could not be represented in a suitable way in a scatterplot. An investigation of researchers' approaches to this topic revealed that word frequency is actually seen as a logarithmic effect, and therefore, word frequency is often transformed into a log(10) value. Consequently, the data from SUBLTEX$_{US}$ was transformed to a Zipf-scale (Heuven et al. 2014) using the following formula: log10 (SUBTLEX$_{US}$ word frequency per million * 1000). With this transformation, the frequencies move closer together, with low frequencies being raised and high frequencies lowered. This makes it much easier to graphically represent the results in an appropriate way.

A by-items as well as a by-subjects analysis was performed on the data (answers > 1 sec were deducted, and the items *ere* and *baik* were not considered). For PHs, the by-items analysis did not show a correlation between log10 word frequency and naming time ($p = 0.13$, $\rho = -0.31$), but a definite trend for PHs with frequent base-words to be pronounced faster. Interestingly, the same cannot be said in the case of real words: the correlation coefficient was close to zero ($p = .61$, $\rho = .11$). The scatterplot shows a cloud with an almost even trend line, thus there is no visible effect of word frequency on naming time.

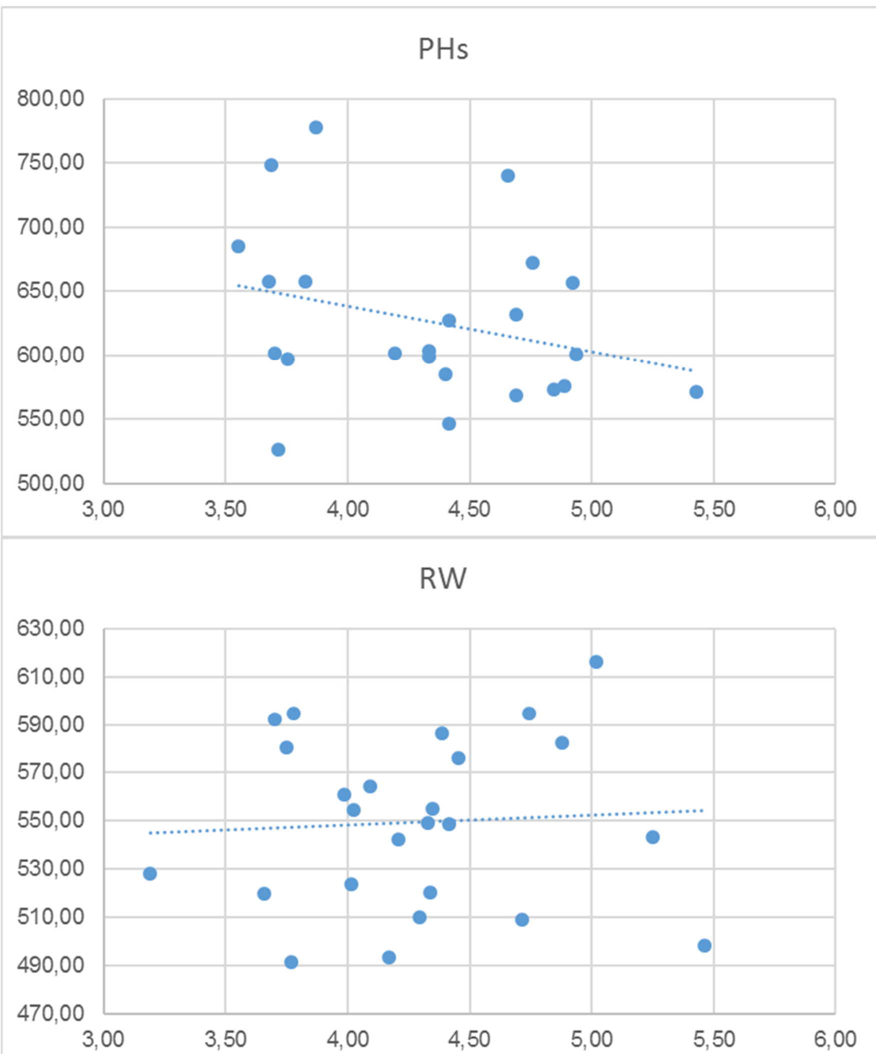


Figure 9 Scatterplots of naming times and log10 (base) word frequency

Figure 9 shows the scatterplots for PHs and real words. To get a clearer picture, a by-subjects analysis was carried out as well. For each participant, his or her naming times were correlated with log10 word frequency, and a one-sample t-test was then performed for all correlation coefficients (N=20). There was no correlation neither for PHs ($t(19) = -1.531$, $p = .14$) nor for real words ($t(19) = 1.519$, $p = .15$). However, the p-values of a few participants reached significance (for 4 participants in the case of PHs and 2 participants in the case of real words, with another 2 participants almost achieving a significant result).

The results of this frequency analysis raise an important question: Why is there no frequency effect, not even for real words? The trend of PHs to be pronounced faster appears somewhat plausible, but the absence of an effect for real words was not anticipated. In returning to the literature review, these results can best be explained by the mixed-list research design. A mixed list prevents the effect from occuring, as the participant is constantly switching between word types and cannot enter into a rhythm. In other naming experiments, mixed-list and pure-block designs were only performed with PHs and nonwords. However, a mixed-list design was shown to have a negative influence on the recall of high and low frequency words. In a pure-list design (in which high and low frequency words are presented separately), high-frequency words are recalled faster, whereas in a mixed-list design, this advantage is not present (e.g., Morin et al. 2006). So if there is no reliable effect of word frequency for real words in mixed lists, it follows that PHs can likewise not be expected to show this effect. In light of this, it is not surprising that the majority of studies presented in Table 1 of this thesis did not find a frequency effect for PHs in a mixed-list design.

Even though there was no significant correlation between word frequency and naming time (cf., Figure 9), there was a trend for PHs with frequent base-words to be pronounced faster, and this trend begs an explanation. On the one hand, PHs take longer to name than real words, and it is a more difficult task for the participant to pronounce them. Possibly, familiar structures aid the process, and contribute to slightly faster naming times (cf., Martin 1982).

c. After reading aloud, do participants remember more real words or PHs?

The basis of this research question is an unexpected sub-task that required participants to name the words and PHs that they remembered after reading aloud the list of stimuli. DMDX presented the stimuli to the participants in random order; to be more precise, the order of items was different for everyone. Participants were only instructed to read the items aloud as quickly as possible; they were not asked to remember any words at all. But despite the unexpected request to recall words, no one objected to the task or complained about it in the least.

Interestingly, results show that PHs are remembered with higher frequency than real words. On average, participants (N=20) remembered 5.3 PHs and 2.3 real words. The total number of remembered words differed greatly among participants and ranged from 3 – 14 words (mean = 7.6). Words that were remembered with high frequency were *bair* (the top answer), *chare*, *focks*, *sord*, *aks*, *ere*, *bread*, *beer*, *owl*, and *skull*. These results are represented graphically in Figure 10.

Frequency did not seem to play a key role, as the highest frequency words (*fone*, *face*, *book*) were rarely remembered. The most frequent answer, *bair*, has a mid-high frequency of 57.41 in the SUBTLEX$_{US}$ database. *Chare* has a similar base-word frequency (49.24), whereas *sord*, *focks*, and *ere* show frequencies that are a bit lower (26.18, 21.61, and 32, respectively). *Aks* has a low frequency (4.88), but might have been remembered due to the conflict that was present during its pronunciation. *Beer* has a fairly high frequency of 75.49, whereas *bread*, *skull* and *owl* show lower frequencies (28.33, 14.71, and 5.61).

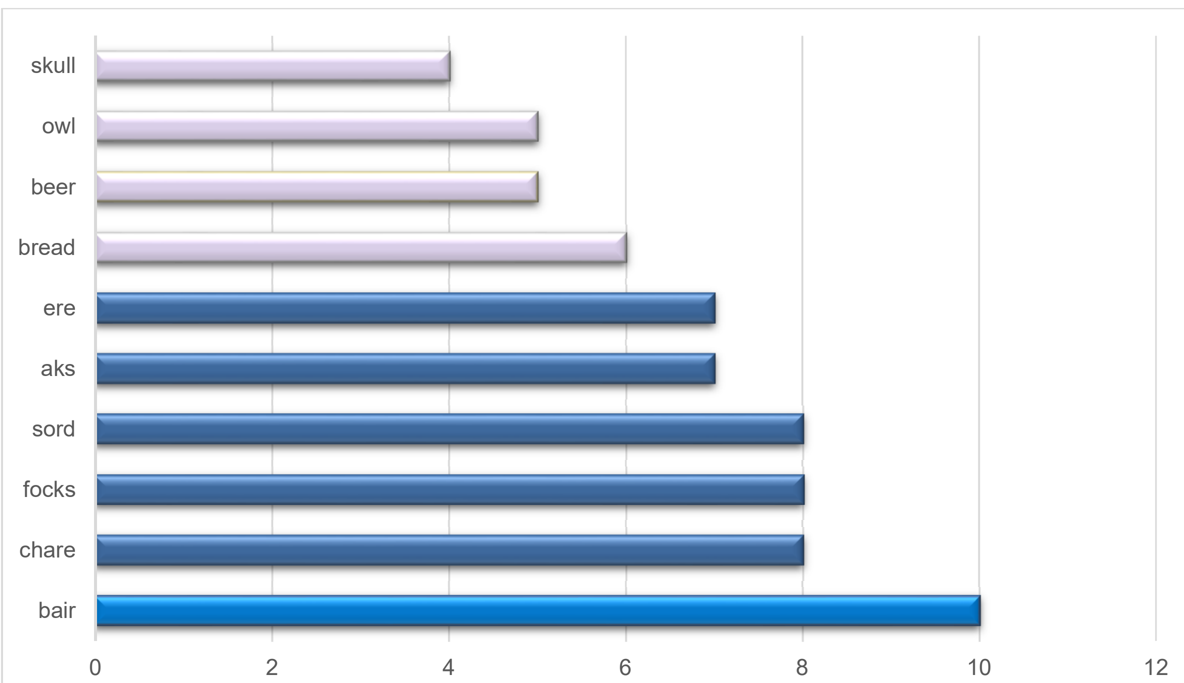


Figure 10 Most frequently remembered words

Responses were sometimes vague, referring to categories such as ‘food’ or ‘animals’. Some responses referred to words that were presented in the practice trial; these responses could not be considered, but showed a similar pattern with PHs being named frequently. Other responses seemed to be memorized in pairs, such as *garlic* and *bread*. In some cases, personal interests might have played a role (whether a person was playing the guitar or if s/he liked owls). It is interesting to note that participants sometimes mispronounced a word (e.g., /bɛrd/ instead of /bɜ:^{r}d/), but later named the animal *bird* among the words that they remembered. Probably, the word was assembled in a different way, but the intended base-word was (co-)activated nonetheless. Seeing that participants remembered about twice as many PHs as real words, the initial hypothesis (that people would remember more real words) was shown to be inaccurate. Possibly, this surprising finding may be attributed to the higher processing effort needed to assemble these unfamiliar words. More attention was paid to the more challenging task of reading aloud strange letter combinations, causing these words to stick in the mind.

The finding that *bair* (bear) was remembered most often suggests that it might be more imageable than the other items in the experiment. However, all included items were highly imageable, and *bear* is not the most imageable of all words with a rating of 572. Actually, *sunset* and *river* have the highest rating (633) according to the MRC Psycholinguistic Database. Returning to Pexman (2012), it should be pointed out that the word 'bear' is semantically rich and has a lot of words associated with it, e.g., in terms of color or strength. It is also a word that children acquire very early in life (because most of them possess a teddy bear), which might be hypothesized to play a role.

7.3 Forced-choice picture task

Following the results of the naming task, the analysis of the three research questions pertaining to the picture task will be presented. These questions investigate the difference in response time between real words and PHs, the difference in the number of correctly identified pictures between real words and PHs, and the difference in response time between the conditions 'semantically related' and 'semantically unrelated', both for real words and PHs.

d. *Is there a difference in response time between real words and PHs?*

In the analysis of response time, the item *baik* was disregarded completely because of its problematic nature. For the item *ere*, responses to the picture were fairly regular with only two mistakes in total; therefore, only the response time of these two instances was deducted. Considering these adjustments, the mean reaction time in the picture task amounted to 737.67 ms for PHs and 710.46 ms for real words. This difference of 27.30 msec is markedly small – it seems that the decision in response to a PH was made with almost the same speed as the decision in response to a real word.

The vast majority of wrong answers was given in response to the item *hownd* (by 10 out of 19 participants), and as a result, this item produced very long response times. Participants had trouble deciding whether a wolf qualified as a hound, and this decision process took time. Therefore, a second analysis was performed in which all responses to this one item were deducted. This leaves a mean response time of 725.03 msec to PHs, which brings the results of the two word groups – real words and PHs – even closer together. Overall, the results suggest that participants have little difficulty in establishing the meaning of PHs, as their decision process is almost equally fast for both word types.

e. *Is there a difference in the number of correctly identified pictures between real words and PHs?*

In the picture task, it was assessed if participants classified a pair (either real word / picture or PH / picture) as a match or mismatch. If a pair was a match (e.g., 'chare' followed by the picture of a chair), but was classified as a mismatch, this was considered an error. All pairs are illustrated in the Appendix to this thesis.

Overall, the picture task worked surprisingly well for the PH stimuli, considering that they appeared on the screen only for 2 seconds each. 95.18 % of all PH / picture combinations were answered correctly, as opposed to 97.68 % of all real word / picture combinations. Errors were few in general (0 - 5 per participant, median = 2, mode = 1). Errors in response to PHs were almost twice as frequent with a mean of 1.16 per participant, as opposed to a mean of 0.58 in response to real words. For both word types, errors were highest in response to semantically related stimuli. Most errors occurred for the following pairs (N=19):

- the word *hownd* combined with the picture of a wolf (10 errors)
- the word *aks* followed by the picture of a hammer (3 errors)
- the word *fruit* followed by the picture of a carrot (3 errors)

The pair *egg* / chicken showed 2 errors in total, as did the stimuli *ere* and *berd*. In the case of the latter two items, the pictures were likely identified as not matching the word because the participants' pronunciation did not match the intended one. In the case of fruit, there might have been confusion as to what exactly qualifies as (a) fruit. A carrot does not belong to the category fruit, and it is not *a* fruit either because it does not have seeds.

Errors on the six semantically unrelated pairs were equally distributed between word groups (0.11 errors per participant for both PHs and real words). Errors on the six semantically related pairs were more frequent following PHs (0.74 vs. 0.37), owing mostly to the problematic item *hownd*.

f. Is there a difference in response time between the conditions 'semantically related' and 'semantically unrelated', both for real words and PHs?

It is perhaps not surprising that participants took longer to respond to PHs if they were paired with semantically related pictures (e.g., *soope* / pizza or *aks* / hammer). The six pairings were arguably a bit more difficult than those in the real word group, and based on the literature review, it could be argued that PHs received less semantic activation because of the orthographic mismatch. The average response time per participant amounted to 894.37 msec for PHs as opposed to 825.96 msec in response to real words. In general, responses to semantically related pictures were above the average response time for all 25 items of a group (156.61 msec and 115.5 per item, respectively), and thus the presence of a conflict is evident. As already mentioned, the item *hownd* had particularly long response times; if this item were deducted, the difference between the groups of words would be a little less pronounced (867.14 vs. 825.96 msec, a difference of 41.18 msec). Still, it seems that PHs are more difficult to classify if they are followed by a picture of an object that belongs to a similar category, as in the example *aks* / hammer. A possible reason might be that the semantic features of their base-words are not activated as strongly as in the case of real words because no orthographic input is given.

For semantically unrelated pairings (e.g., *river* / bride or *bench* / goat), the analysis of response times shows similar results. Responses given in response to real words were 22.25 msec faster on average (736.72 vs. 714.47 msec). This difference is fairly small, and when taking a look at the individual results of the participants, it turns out that some of them gave a faster response following the presentation of a real word, whereas for others, the opposite was the case. Still, the difference between the two word groups is much less pronounced than in the case of semantically related words. Again, there is a conflict between the word and picture stimuli, but one that can apparently be resolved much more easily. It would therefore appear that due to the simpler nature of this task, it makes little difference if the preceding stimulus is a real word or a PH.

After taking part in the experiment, some participants made interesting comments about the word-picture combinations in the semantically related condition. Essentially, they were trying to explain why they had answered 'yes' even though the picture was actually not a match, which they only realized too late. The word *owl* was paired with the picture of an eagle, and one participant commented “I expected to see a picture of a bird and there it was”. The word *hedge* was followed by the picture of a tree, and 2 persons said “A tree is reasonably similar to a hedge”. The greatest confusion was caused by the pair *hownd* / wolf, which caused many participants to say “A wolf and a hound is pretty close, so I thought ‘well, that's good enough’”. Regarding the pair *sord* / shield, 2 persons said “Sword and shield seemed very similar to me”. All of these pairs are similar by design, but the pair *hownd* / wolf was arguably too similar and resulted in the highest number of mistakes.

8. Discussion

The task given to the participants was a bit out of the ordinary: half of the words were spelled in a strange way, and many differed from their base-word to such an extent that they were barely recognizable, or maybe not at all. In some cases, the items were in fact more similar in spelling to other words, e.g., *angker* was spelled like *anger* with the exception of one additional letter, and *aks* was spelled like *ask* with two mixed-up letters. This similarity was unintentional, as the items had been selected from prior studies involving PHs. From a look at the data, it is clear that this visual similarity was a source of confusion that hindered swift pronunciation of the target word. The opposite was true for words that were visually very similar to their base-words, which was the case for *leef* and *sord*, and these items were in fact pronounced fastest in the PH group. It can therefore be confirmed that visual similarity has a strong effect on naming time (c.f., Martin 1982): on the one hand, an adverse effect if the spelling can lead to a confusion with other words; on the other hand, a favorable effect if the spelling leads to fast identification of the target word.

Despite these distractions, it was clear that the participants were aware of the meaning of the base-words. After having finished a task, some would say they remembered *bear* but spelled with ‘ai’ (to give only one example), or they would state that they had been aware of two possible pronunciations, such as *bike* and *bake*. No one had been instructed to guess what was meant by the strange spelling, much less to remember a word, and intervals between items had been quite short. Even if the pronunciation of a word was a bit off, its base-word was still recognized, which was confirmed by it being named among the remembered words, and/or by giving the correct answer to the picture stimulus in the subsequent picture task. It is conceivable that a PH or nonword activates a multitude of information that is stored in the brain, and as a result, fictional words can trigger semantic information, however small this activation may be.

In general, the tasks were received well by the participants. They considered the tasks to be manageable, and in many cases, even fun. The images were generally perceived as natural and suitable, but it still seems that there is no one prototypical image that fits everyone's mental image. One problem that presented itself was that many test-takers appeared to see the tasks as a kind of game; they later said that they were trying to employ a 'strategy' such as visualizing the object or animal when reading the word aloud in order to be able to give a faster answer. Such a strategy may well have had an effect on the data. In the naming task, participants were often a bit hesitant and were trying to make sense of the unfamiliar letters; as a result, they did not read the words aloud as quickly as possible. Typical times for reading aloud were in the vicinity of 500 ms (although there are individual differences), anything closer to or above one second was likely due to hesitation or some kind of confusion. One participant pointed out that *aks* was actually common for *ask* in African American Vernacular English, as well as other dialects. This is something that I had not been aware of, and that seemed to generate confusion among a few participants. It is perfectly conceivable that hesitation and confounding factors play a role in similar experiments as well, making it difficult to compare the reading of PHs to nonword reading and particularly to the reading of real words. Still, it should be stated that there was surprising consistency among native speakers of English in the pronunciation of many unfamiliar words. This consistency is clearly an argument in favor of the existence of a rule-based GPC system.

Another factor that clearly influenced the data was knowledge of a foreign language. When conducting a similar experiment in an English-speaking country, especially in the United States, it may well be the case that the participants do not speak any second language, and thus their way of dealing with unfamiliar letter strings would be completely unbiased. By contrast, almost every native speaker of English who took part in this experiment knew at least one other language, if not more. Only those few who were in Vienna on vacation or on a business trip did not know a foreign language. Those residing in Austria on a long-term basis are likely to speak German quite well; three participants were even former or current students of German. It was interesting to see that those with excellent knowledge of German tended to pronounce words differently than others, e.g., they pronounced *soope* as 'soap' instead of 'soup' and *baik* as 'bike' instead of 'bake'. All stimuli used in the present experiment had been taken from other naming experiments, but unfortunately, academic papers of this sort do not usually include a phonetic transcription of the target word. As a result, two target words were assumed to have a different pronunciation than they actually did when they were pronounced by a native speaker.

The letter combination *baik* is in fact pronounced 'bake' by the vast majority, and *ere* is pronounced like 'air' by many native speakers (or like 'ere' as in an archaic version of 'before'). To a non-native speaker, the target pronunciation may sometimes only reveal itself way into the research process. The finding that participants with excellent knowledge of German pronounced words differently is certainly highly interesting. It suggests that our GPC system is not static, but changes over time due to external influences. While it would have been much easier to conduct this kind of experiment at an American university, with a homogeneous group of monolinguals, many interesting realizations would have been lost. As a follow-up project, it would be interesting to investigate the differences between native and non-native speakers of English with regard to specific nonwords.

The most interesting finding of this study is probably the high number of PHs that participants remembered after the naming task. They had not received any instructions to that end, hence they could not have employed any strategy to perform well on this unexpected 'sub-task'. Indeed, instructions had been minimal, with an emphasis on swift reading aloud. It seems likely that the effort in assembling a pronunciation for unfamiliar words caused these words to stick in the mind, reflecting a higher processing effort. Most participants pronounced one or more words unlike the intended target, but when asked which words they remembered, they named the target word with its accurate pronunciation. This finding indicates that the meaning of the underlying base-words had been understood by the participant; otherwise it would have been hard to remember a word solely by recalling its pronunciation. Some kind of activation must have taken place, even though there were no instructions that would have necessitated the activation of meaning. This activation appears to occur involuntarily, as a byproduct of reading, following phonological mediation. Such a 'spreading activation' from pronunciation to meaning supports the petal metaphor presented earlier in this paper. The word's pronunciation is computed, causing the stored sound pattern to be activated, and along with this activation, other kinds of information (smell, taste, visual characteristics, connotation, associated conception…) are activated as well. This idea is illustrated graphically in Figure 11.

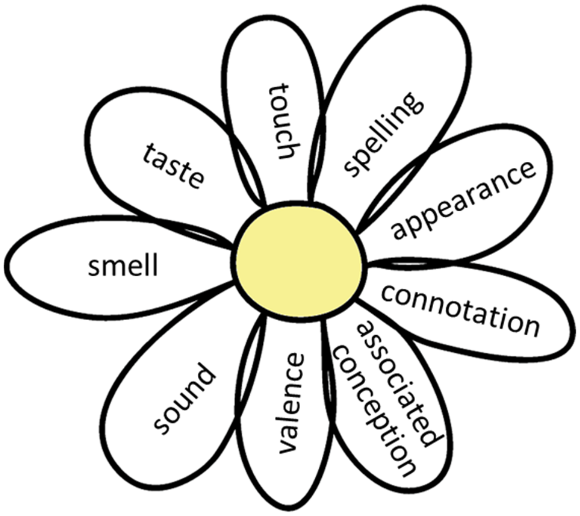


Figure 11 The petal metaphor (based on flower petals template, Clipart Library 2016)

Clearly, a word is more than just a label that we give to a thing. We may recognize a word either through its spelling or its sound pattern (or even through visual or tactile information in the case of sign language or braille), and in our minds, these labels are linked to a whole array of information. A word's connotation or its associated conception (Frege 1948) may come to mind; we may remember a positive or negative experience connected with the overall concept. We may see it in front of our inner eye, or we may think of its smell, or about the way that it tastes, or about what it feels like to touch it. Not all words will have an equal number of petals; there will be richer and poorer representations depending on our experience with the word in question. Any petal, i.e., any aspect of meaning, may cause other aspects to be activated. A smell can remind us of a fruit or an animal; a color can remind us of a particular object. All these aspects are interconnected, and it therefore makes little sense to view them in isolation.

Yet this is what studies of naming and models of reading aloud typically do: spelling translates into sound, other aspects are left aside. At an early stage of investigation, this approach is understandable, but after decades of research into the topic, models have advanced to a point where a wholesome model needs to be developed. As discussed at length in section 3.3, PDP is the only model that takes semantics into account, whereas DRC and CDP+ only include a sort of place holder in their architecture. However, the dual-route architecture of DRC and CDP+ is favored by many research findings, e.g., by the realization that nonword reading is largely preserved in patients suffering from semantic dementia (cf., Graham, Patterson & Hodges 2000: 157). This finding indicates that words and nonwords are processed differently; if one type of word can still be read whereas reading of the other type is severely impaired, there is no other explanation. The hybrid model CDP+ is very promising insofar as it combines a dual-route architecture and localist representations with a connectionist system for nonword reading. In recent years, CDP+ has been extended to read more than one syllable, turning it into CDP++ (Perry, Ziegler & Zorzi 2010), and it has been applied to other languages besides English. While these are doubtlessly important extensions, it would be desirable to begin work on a wholesome model that accounts for the activation of meaning in the reading of real words, PHs, and partially nonwords as well. Such an implementation will certainly not be realizable with the same mathematical precision as other ventures, but will be of great scientific value because it provides a true-to-life depiction of what is going on in our minds when we read.

Notably, one word was remembered with much higher frequency than all other stimuli. The PH *bair* was recalled by 10 out of 20 participants – a result that suggests that there might be something special about this particular item. It is short, but not one of the shortest items in the experiment. It is fairly frequent, but nowhere near the most frequent items. As regards orthographic similarity between the PH and its base-word, it may be observed that *bair* and *bear* do not look very similar. This is interesting because Harm & Seidenberg (2004: 706-707) argue that in their model, semantic activation stems partly from orthographic similarity to real words. And while this may be the case, it seems that other mechanisms play a stronger role in triggering semantic activation. It is likely that *bair* was remembered with high frequency due to the rich semantic representation of the base-word *bear* (a large number of petals are activated), or else due to acquisition of the word at an early age. The outstanding result of *bair* stands in contrast to a number of items that were not recalled once. In fact, *noze* was the only PH that was never recalled; among real words, this happened quite frequently (in the case of *sunset*, *hedge*, *fruit*, *orange*, and *belt*). In any case, the finding that only one PH was not remembered by any participant clearly shows that semantic activation takes place to a large extent for this word group.

9. Conclusion

This thesis set out to investigate the role of semantics in reading aloud. To be more specific, the object of investigation was the activation of semantics in the reading of curiously spelled words. Do people associate meaning with such funny, unfamiliar letter combinations? Or are they just an alien structure that poses an obstacle to reading?

As was discussed at length in the theoretical part of this paper, two major models of reading aloud are based on a dual-route architecture, whereas a third model assumes that all words are read via the same mechanism. While both approaches have their pros and cons, two findings that were replicated as a part of this study support the existence of two distinct routes for reading aloud: the surprising consistency among native speakers in the pronunciation of unfamiliar words, and the difference in naming time between real and made-up words. It makes sense that words that are known to the reader, and that are encountered with high frequency, no longer need to be assembled one letter at a time, and their spelling can be accessed as a whole. If these fantasy words were encountered often enough, they might themselves receive an entry in a hypothetical mental lexicon, and could be retrieved directly. It appears that as schoolchildren, we learn to associate letters with known phonological patterns, and later learn to recognize frequent spellings as a whole to speed up reading. Still, the original mechanism remains, or we would have no way to process unfamiliar or lesser known words. This mechanism is what Coltheart et al. (2001) refer to as GPC rules. Moreover, Graham, Patterson & Hodges (2000) showed that patients suffering from semantic dementia have less difficulty in reading nonwords than real words, which is clearly an argument in favor of two distinct processing routes.

As suggested by the data collected in this experiment, these GPC rules are not entirely static, but may change over time due to external influences. Those participants who spoke German very well because they had either studied the language at university or lived in Austria for a long time, showed a tendency to pronounce certain words differently than their fellow native speakers of English. Clearly, they must have developed separate rules for pronouncing German words, and these rules are interfering with the rules that they are applying to process their native language. As a follow-up study, it would be interesting to compare English and German native speakers with regard to their reading of low-frequency or made-up words.

Interestingly, no frequency effect was found in the present study, neither for real words nor for PHs. The latter group showed a tendency for items with frequent base-words to be read faster, but the regression analysis did not reach significance. In the literature review, a number of studies were discussed that investigated a frequency effect in mixed lists (McCann & Besner 1987; Herdman, LeFevre & Greenham 1996; Grainger, Spinelli & Ferrand 2000); the majority did not find an effect in this experiment design, but none of these studies used real words as stimuli. That notwithstanding, it appears that a mixed-list design simply does not allow for an effect of word frequency to occur, regardless of the nature of the stimuli. The distraction appears to prevent participants from entering into some kind of rhythm, and as a result, visual similarity to the base-word (as in *leef* / leaf and *sord* / sword, cf. Martin 1982) plays a far greater role.

However, a possible frequency effect in pure lists could not be confirmed, as people were not tested in a pure-list design.

A particularly striking finding was that participants remembered PHs about twice as frequently as real words. They had not been instructed to remember any words at all, but still, they listed quite a few, and it appeared that the funnily spelled words had stuck in their minds. Even if their pronunciation of the word had been a little bit off, they clearly associated it with the meaning of its base-word, or stated that they had been aware of two conflicting possibilities. So somewhere along the reading process, semantic information is accessed, even if the task does not necessitate any activation to that end. Rather, it is an involuntary byproduct of the reading process, and it seems likely that even nonwords activate some kind of information, especially if they are written or sound like an existing word. It was observed that PHs were remembered easily (possibly due to the greater effort that was required in their assembly), but given that the decision process for semantically related pairs took quite long for PHs, it is unlikely that these stimuli receive greater semantic activation than real words. In the naming task, *bair* was by far the most frequent answer, possibly because of its rich semantic representation. The automatic activation of meaning strongly calls for the development of a wholesome model of reading aloud that accounts for all aspects of reading: orthography, phonology, and semantics. To date, PDP is the only such model, and dual-route models have yet to follow its example.

In returning to the overall research question of this thesis, I am inclined to conclude that meaning is activated to a large extent during PH reading. When reading about experiment designs involving PHs, one would think that study participants do not associate such funny words with their base-words when reading them, but this does not seem to be the case, at least not with highly imageable words. Instead, the results of the experiment presented in section 7 suggest that unfamiliar spellings do not hinder activation; on the contrary, the increased effort involved in the processing of these curiously spelled words appears to be the reason why participants remembered them about twice as often as regularly spelled words. While it is difficult to say what exact aspects were activated, it was clear in talking to the participants that the base-words of the PHs had been recognized. A naming task would not necessarily require an understanding of what is being read, but still, the mind seems to be actively engaged in meaning-making during such a task. This process occurred via phonology; the phonological representation was computed first and led to the activation of other aspects. Phonology is acquired first in language acquisition, so it could be hypothesized that it is a strong trigger of meaning. In the picture task, it was shown that participants had little if any difficulty in deciding if the picture matched the word stimulus, not even in the semantically related condition (which can be quite confusing even with real words). Taking dual-route models as a basis, the strong memory effect for PHs might mean that assembled words are read slower, but lead to a better recall of the word after reading. In today's classrooms, children are constantly pushed to read faster, but considering the evidence presented here, faster reading may not necessarily be the desired goal.

In conclusion, it appears that various aspects of a word may be accessed as part of the reading process: the spelling (of the word itself and of visually similar words), the word's pronunciation (and perhaps conflicting pronunciations and word neighbors), the assumed sense(s) of the word, its semantic features (which are linked to sensory information), its connotation and associated conception (Frege 1948). This idea was expressed by means of the petal metaphor, which nicely shows that all of these aspects are interconnected. Orthography and phonology are only two aspects, and while these are the ones which are easiest to represent in a computational model, they show only a fragment of what there is to see. A flower with only two petals would hardly look like a flower, but with six or more petals, we may view it in its entire beauty.

10. References

Ahlsén, Elizabeth. 2006. *An introduction to neurolinguistics*. Amsterdam: John Benjamins.

Aitchison, Jean. 2012. *Words in the mind*. 4th edition. Chichester: Wiley-Blackwell.

Baron, Jonathan. 1977. "Mechanisms for pronouncing printed words: use and acquisition". In: LaBerge, David; Samuels, S. Jay (eds.). *Basic processes in reading: perception and comprehension*. Hillsdale, N.J.: Erlbaum, 175-216.

Beech, John R. 2010. "Young readers' strategic approaches to reading unfamiliar words in text". *Reading & Writing Quarterly* 26, 264-284.

Besner, Derek; Twilley, Leslie; McCann, Robert S.; Seergobin, Ken; Kintsch, Walter. 1990. "On the association between connectionism and data: are a few words necessary?". *Psychological Review* 97(3), 432-446.

Besner, Derek; Davelaar, Eileen. 1983. "Suedohomofoan effects in visual word recognition: evidence for phonological processing". *Canadian Journal of Psychology* 37(2), 300-305.

Borowsky, Ron; Owen, William J.; Masson, Michael E.J. 2002. "Diagnostics of phonological lexical processing: pseudohomophone naming advantages, disadvantages, and base-word frequency effects". *Memory & Cognition* 30(6), 969-987.

Borowsky, Ron; Masson, Michael E.J. 1999. "Semantic ambiguity effects in word identification". *Journal of Experimental Psychology, Learning, Memory & Cognition* 22, 63-85.

Braun, Mario; Hutzler, Florian; Münte, Thomas F.; Rotte, Michael; Dambacher, Michael; Richlan, Fabio; Jacobs, Arthur M. 2015. "The neural bases of the pseudohomophone effect: phonological constraints on lexico-semantic access in reading". *Neuroscience* 295, 151–163.

Brown, Roger; McNeill, David. 1966. "The 'tip of the tongue' phenomenon". *Journal of Verbal Learning and Verbal Behavior* 5(4): 325-337.

Brysbaert, Marc; New, Boris. 2009. "Moving beyond Kučera and Francis: a critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English". *Behavior Research Methods* 41(4), 977-990.

Busemeyer, Jerome R.; Diederich, Adele. 2010. *Cognitive Modeling*. Los Angeles: Sage.

Carroll, Lewis. 1998. *Alice's adventures in wonderland and Through the looking glass and what Alice found there*. London: Penguin Books.

Carroll, Lewis. 1872. *Through the looking glass, and what Alice found there*. (ed. by Hugh Haughton). London: Macmillan.

Carroll, Lewis. 1865. *Alice's adventures in wonderland*. London: Macmillan.

Cembrani, Veronica. 2010. *Cascaded and thresholded processing in visual word recognition: does the Dual Route Cascaded model require a threshold?* PhD thesis, University of Trento.

Coltheart, Max. 2012. "Dual-route theories of reading aloud". In Adelman, James S. (ed.). *Visual word recognition volume 1: models and methods, orthography, and phonology*. London: Psychology Press, 3-27.

Coltheart, Max. 2005. "Modeling reading: the dual-route approach". In: Snowling, Margaret J.; Hulme; Charles (eds.). *The science of reading: a handbook*. Oxford: Blackwell, 6-23.

Coltheart, Max; Rastle, Kathleen; Perry, Conrad; Langdon, Robyn; Ziegler, Johannes. 2001. "DRC: a dual route cascaded model of visual word recognition and reading aloud". *Psychological Review,* 108(1), 204-256.

Coltheart, Max; Curtis, Brent; Atkins, Paul; Haller, Michael. 1993. "Models of reading aloud: dual-route and parallel-distributed-processing approaches. *Psychological Review* 100(4), 589-608.
Coltheart, Max. 1981. "The MRC psycholinguistic database". *Quarterly Journal of Experimental Psychology* 33A, 497-505.
Coltheart, Max. 1978. "Lexical access in simple reading tasks". In Underwood, Geoffrey (ed.). *Strategies of information processing*. London: Academic Press, 151-216.
Darcy, Isabelle. 2010. "DMDX Introductory Tutorial". Indiana University Bloomington. http://www.iub.edu/~psyling/resources/dmdx-tutorial_id_2010.pdf (15 Jun. 2017).
Davis, Matt. 1999. "DMDX Introductory Tutorial". University of Cambridge. http://www.mrc-cbu.cam.ac.uk/personal/matt.davis/dmdx-tutorial/dmdx-tutorial.ppt (15 Jun. 2017).
Eysenck, Michael W.; Keane, Mark T. 2010. *Cognitive psychology: a student's handbook*. 6th edition. Hove: Psychology Press.
Forster, Kenneth I.; Forster, Jonathan C. 2003. "DMDX: a windows display program with millisecond accuracy". *Behavior Research Methods, Instruments, & Computers*, 35(1), 116-124.
Forster, Kenneth I.; Chambers, Susan M. 1973. "Lexical access and naming time". *Journal of Verbal Learning and Verbal Behavior* 12, 627-635.
"Free Flower Templates Printable #1319314". 2016. *Clipart Library*. http://clipart-library.com/clipart/8TxnedaRc.htm (accessed 10 Feb. 2018).
Frege, Gottlob. 1948. "Sense and reference". *The Philosophical Review* 57(3), 209-230.
Fromkin, Victoria; Rodman, Robert; Hyams, Nina. 2003. *An introduction to language*. 7th edition. Boston, MA: Wadsworth.
Gleason, Jean B. 1958. "The child's learning of English morphology". *Word* 14(2-3), 150-177.
Glushko, Robert J. 1979. "The organization and activation of orthographic knowledge in reading aloud". *Journal of Experimental Psychology: Human Perception and Performance* 5(4), 674-691.
Graham, Naida L.; Patterson, Karalyn; Hodges, John R. 2000. "The impact of semantic memory impairment on spelling: evidence from semantic dementia". *Neuropsychologia* 38(2), 143-163.
Grainger, Jonathan; Spinelli, Elsa; Ferrand, Ludovic. 2000. "Effects of base-word frequency and orthographic neighborhood size in pseudohomophone naming". *Journal of Memory and Language* 42, 88-102.
Harm, Michael W.; Seidenberg, Mark S. 2004. "Computing the meanings of words in reading: cooperative division of labor between visual and phonological processes". *Psychological Review* 111(3), 662-720.
Harm, Michael W. 2002. "Building large scale distributed semantic feature sets with WordNet". Technical Report. Pittsburgh, PA: Carnegie Mellon University, Center for the Neural Basis of Cognition.
Harm, Michael W.; Seidenberg, Mark S. 1999. "Phonology, reading acquisition, and dyslexia: insights from connectionist models". *Psychological Review* 106(3), 491-528.
Herdman, Chris M.; LeFevre, Jo-Anne; Greenham, Stephanie L. 1996. "Base-word frequency and pseudohomophone naming". *The Quarterly Journal of Experimental Psychology* 49A(4), 1044-1061.
Heuven, Walter J.B. van; Mandera, Pawel; Keuleers, Emmanuel; Brysbaert, Marc. 2014. Subtlex-UK: a new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology* 67, 1176-1190.

Houghton, George; Zorzi, Marco. 2003. “Normal and impaired spelling in a connectionist dual-route architecture”. *Cognitive Neuropsychology* 20, 115-162.
Jansen, Peter; Watter, Scott. 2008. “SayWhen: An automated method for high-accuracy speech onset detection”. *Behavior Research Methods* 40(3), 744-751.
Jared, Debra. 2002. “Spelling-sound consistency and regularity effects in word naming”. *Journal of Memory and Language* 46(4), 723-750.
Kessler, Brett; Treiman, Rebecca. 2002. “Phonetic biases in voice key response time measurement”. *Journal of Memory and Language* 47, 145-171.
Marian, Viorica; Bartolotti, James; Chabal, Sarah; Shook, Anthony. 2012. “CLEARPOND: Cross-linguistic easy-access resource for phonological and orthographic neighborhood densities”. PLoS ONE 7(8): e43230. https://doi.org/10.1371/journal.pone.0043230 (15 Jun. 2017).
Marmurek, Harvey H.C.; Kwantes, Peter J. 1996. “Reading words and wirds: phonology and lexical access”. *The Quarterly Journal of Experimental Psychology* 49A(3), 696-714.
Martin, Randi C. 1982. “The pseudohomophone effect: the role of visual similarity in non-word decisions”. *The Quarterly Journal of Experimental Psychology Section A Human Experimental Psychology* 34(3), 395-409.
McCann, Robert S.; Besner, Derek; Davelaar, Eileen. 1988. “Word recognition and identifcation: Do word frequency effects reflect lexical access?“ *Journal of Experimental Psychology: Learning, Memory, and Cognition* 14(4), 693-706.
McCann, Robert S.; Besner, Derek. 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(1), 14-24.
McClelland, James L.; Rumelhart, David E. 1981. “An interactive activation model of context effects in letter perception: part 1. An account of basic findings”. *Psychological Review* 88(5), 375-407.
McKay, Adam; Davis, Chris; Savage, Greg; Castles, Anne. 2008. “Semantic involvement in reading aloud: evidence from a nonword training study”. *Journal of Experimental Psychology: Learning, Memory and Cognition* 34(6), 1495-1517.
Meara, Paul; Miralpeix, Imma. 2015. “V_YesNo”. http://www.lognostics.co.uk/tools/V_YesNo/V_YesNo.htm (accessed 15 May 2017).
Miller, George. 1998. “Nouns in WordNet”. In Fellbaum, Christiane (ed.). *WordNet: An electronic lexical database*. Cambridge, MA: MIT Press, 23-46.
Miller, George. 1995. “WordNet: a lexical database for English”. *Communications of the ACM* 38(11), 39-41.
Morin, Carolin; Poirier, Marie; Fortin, Claudette; Hulme, Charles. 2006. “Word frequency and the mixed-list paradox in immediate and delayed serial recall”. *Psychonomic Bulletin and Review* 13(4), 724-9.
Morton, John. 1969. “The interaction of information in word recognition”. *Psychological Review* 76(2), 165–178.
“non-word, n.”. September 2016. *OED* Online. Oxford: Oxford University Press. http://www.oed.com.uaccess.univie.ac.at/view/Entry/128167?redirectedFrom=non-word#eid (accessed 10 Jul. 2017).
Papadatou-Pastou, Marietta. 2011. “Are connectionist models neurally plausible? A critical appraisal”. *Encephalos: archives of neurology and psychiatry* 48(1), 5-11.
Patterson, Karalyn; Nestor, Peter J.; Rogers, Timothy T. 2007. “Where do you know what you know? The representation of semantic knowledge in the human brain”. *Nature Reviews Neuroscience* (8), 976-987.

Patterson Karalyn; Morton, John. 1985. "From orthography to phonology: an attempt at an old interpretation". In Patterson, Karalyn; Marshall, John C.; Coltheart, Max (eds.). *Surface dyslexia: neuropsychological and cognitive studies of phonological reading*. London: Lawrence Erlbaum Associates, 335-359.
Perry, Conrad; Ziegler, Johannes C.; Zorzi, Marco. 2010. "Beyond single syllables: large-scale modeling of reading aloud with the connectionist dual process (CDP++) model". *Cognitive Psychology* 61(2), 106-151.
Perry, Conrad; Ziegler, Johannes C.; Zorzi, Marco. 2007. "Nested incremental modeling in the development of computational theories: the CDP+ model of reading aloud". *Psychological Review*, 114(2), 273-315.
Pexman, Penny M. 2012. "Meaning-based influences on visual word recognition". In Adelman, James S. (ed.). *Visual word recognition volume 2: meaning and context, individuals, and development*. Hove: Psychology Press, 24–43.
Plaut, David C.; McClelland, James L.; Seidenberg, Mark S.; Patterson, Karalyn. 1996. "Understanding normal and impaired word reading: computational principles in quasi-regular domains". *Psychological Review* 103, 56-115.
Pritchard, Stephen C.; Coltheart, Max; Marinus, Eva; Castles, Anne. 2016. "Modelling the implicit learning of phonological decoding from training on whole-word spellings and pronunciations". *Scientific Studies of Reading* 20(1), 49-63.
Pritchard, Stephen C.; Coltheart, Max; Palethorpe, Sallyanne; Castles, Anne. 2012. "Nonword reading: comparing dual-route cascaded and connectionist dual-process models with human data". *Journal of Experimental Psychology: Human Perception and Performance* 38(5), 1268-1288.
Protopapas, Athanassios. 2007. "CheckVocal: a program to facilitate checking the accuracy and response time of vocal responses from DMDX". *Behavioral Research Methods* 39(4), 859-862.
R Core Team. 2015. *R: a language and environment for statistical computing*. (Version 3.2.3). [Computer Program]. Vienna: R Foundation for Statistical Computing.
Reynolds, Michael; Besner, Derek. 2005. "Basic processes in reading: a critical review of pseudohomophone effects in reading aloud and a new computational account". *Psychonomic Bulletin & Review* 12(4), 622-646.
Richardson, Fiona M.; Seghier, Mohamed L.; Leff, Alexander P.; Thomas, Michael S.C.; Price, Cathy J. 2011. "Multiple routes from occipital to temporal cortices during reading". *Journal of Neuroscience* 31(22), 8239–8247.
Rodd, Jennifer M. 2004. "The effect of semantic ambiguity on reading aloud: a twist in the tale". *Psychonomic Bulletin & Review* 11(3), 440-445.
Rosch, Eleanor. 1975. "Cognitive representations of semantic categories". *Journal of Experimental Psychology* 104(3): 192-233.
Roux, Frédéric; Armstrong, Blair C.; Carreiras, Manuel. 2017. "Chronset: an automated tool for detecting speech onset". *Behavior Research Methods* 49(5), 1864-1881.
Rumelhart, David E.; McClelland, James L.; PDP Research Group. 1986. *Parallel distributed processing: explorations in the microstructure of cognition*. MIT Press: Cambridge.
Rumelhart, David E.; McClelland, James L. 1982. "An interactive activation model of context effects in letter perception: part 2. The contextual enhancement effect and some tests and extensions of the model". *Psychological Review* 89(1), 60-94.
Saussure, Ferdinand de. 1986. *Course in general linguistics*. (ed. by Charles Bally, Albert Sechehaye, with the collaboration of Alfred Riedlinger). Chicago: Open Court.
Seidenberg, Mark. 2007. "Connectionist models of reading". In Gaskell, Gareth (ed.). *Oxford Handbook of Psycholinguistics*. Oxford: Oxford University Press, 235-250.

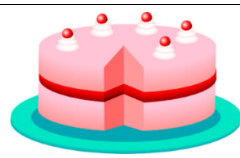
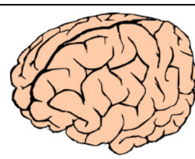
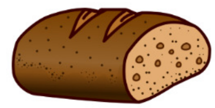
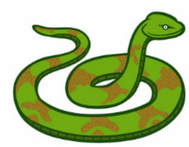
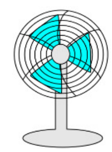
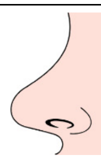
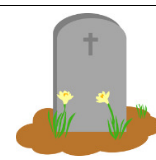
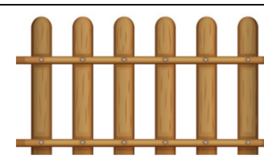
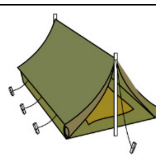
Seidenberg, Mark S.; Plaut, David C. 2006. "Progress in understanding word reading: data fitting versus theory building". In Andrews, Sally (ed.). *From inmarks to ideas: current issues in lexical processing*. London: Psychology Press, 25-49.
Seidenberg, Mark S.; Petersen, Alan; MacDonald, Maryellen C.; Plaut, David C. 1996. "Pseudohomophone effects and models of word recognition". *Journal of Experimental Psychology: Learning, Memory, and Cognition* 22(1), 48-62.
Seidenberg, Mark S.; McClelland, James L.; Kintsch, Walter. 1990. "More words but still no lexicon: reply to Besner et al. (1990)". *Psychological Review* 97(3), 447-452.
Seidenberg, Mark S.; McClelland, James L. 1989. "A distributed, developmental model of word recognition and naming". *Psychological Review*, 96(4), 523-568.
Simos, Panagiotis G.; Breier, Joshua I.; Fletcher, Jack M.; Foorman, Barbara R.; Castillo, Eduardo M.; Papanicolaou, Andrew C. 2002. "Brain mechanisms for reading words and pseudowords: an integrated approach". *Cerebral Cortex* 12, 297-305.
Stone, Gregory O.; Orden, Guy C. van 1993. "Strategic control of processing in word recognition". *Psychology* 19, 744-774.
Strain, Eamon; Patterson, Karalyn; Seidenberg, Mark S. 1995. "Semantic effects in single-word naming". *Journal of Experimental Psychology* 21(5), 1140-1154.
Taft, Marcus; Russell, Bruce. 1992. "Pseudohomophone naming and the word frequency effect". *The Quarterly Journal of Experimental Psychology* 45A(1), 51-71.
Wickelgren, Wayne A. 1969. "Context-sensitive coding, associative memory, and serial order in (speech) behavior". *Psychological Review* 76, 1-15.
Wilson, Michael D. 1988. "The MRC Psycholinguistic Database: machine readable dictionary, Version 2". *Behavioural Research Methods, Instruments and Computers* 20, 6-11.
Woore, Robert. 2010. "Thinking aloud about L2 decoding: an exploration of the strategies used by beginner learners when pronouncing unfamiliar French words". *Language Learning Journal* 38(1), 3-17.
"word, n. and int". September 2016. *OED* Online. Oxford: Oxford University Press. http://www.oed.com.uaccess.univie.ac.at/view/Entry/230192?rskey=V9D3Pq&result=1#eid (accessed 10 Jul. 2017).
Ziegler, Johannes C.; Jacobs, Arthur M.; Klüppel Dorothee. 2001. "Pseudohomophone effects in lexical decision: still a challenge for current word recognition models". *Journal of Experimental Psychology: Human Perception and Performance* 27(3), 547-559.
Zorzi, Marco; Houghton, George; Butterworth, Brian. 1998. "Two routes or just one in reading aloud? A connectionist dual-process model". *Journal of Experimental Psychology* 24(4), 1131-1161.

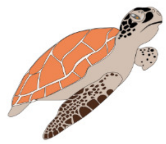
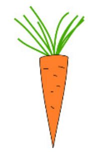
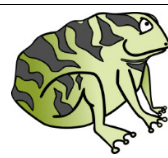
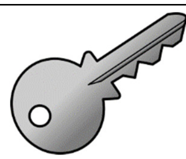
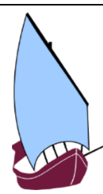
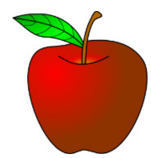
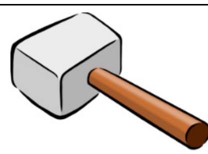
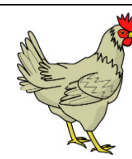
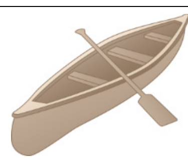
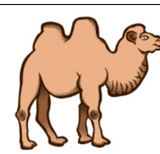
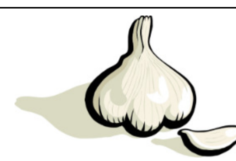
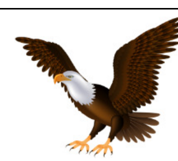
Pictures shown in the Appendix (all accessed Jul./Aug. 2017)

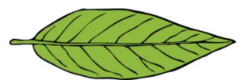
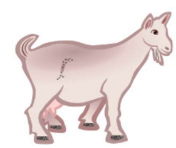
"A coffee cup". 2017. *Open Clipart*. https://openclipart.org/detail/276267/a-coffee-cup.
"Apple". 2009. *Open Clipart*. https://openclipart.org/detail/24319/apple.
"A simple belt". 2010. *Open Clipart*. https://openclipart.org/detail/101503/a-simple-belt.
"Bag". 2007. *Open Clipart*. https://openclipart.org/detail/3333/bag.
"Baking Cliparts #21047". n.d. *Clipart Library*. http://clipart-library.com/clipart/399018.htm.
"Ball". 2010. *Open Clipart*. https://openclipart.org/detail/66751/beach-ball.
"Flag of California Bear". 2012. *Open Clipart*. https://openclipart.org/detail/172279/flag-of-california-bear.
"Book". 2007. *Open Clipart*. https://openclipart.org/detail/9358/book.
"Bomb". 2009. *Open Clipart*. https://openclipart.org/detail/22044/bomb.
"Brain". 2011. *Open Clipart*. https://openclipart.org/detail/140701/brain.

"Bread". 2015. *Open Clipart*. https://openclipart.org/detail/213638/bread.
"Bride 1". 2005. *Open Clipart*. https://openclipart.org/detail/182937/bride-1.
"Camel coloured". 2015. *Open Clipart*. https://openclipart.org/detail/215616/camel-coloured.
"Carrot". 2015. *Open Clipart*. https://openclipart.org/detail/213503/carrot.
"Chain". 2009. *Open Clipart*. https://openclipart.org/detail/22077/chain.
"Carnival Mask PNG". n.d. *Gallery YoPriceVille*. http://gallery.yopriceville.com/Free-Clipart-Pictures/Carnival-Mask-PNG/Purple_Carnival_Mask_PNG_Clip_Art_Image.
"Chicken". 2009. *Open Clipart*. https://openclipart.org/detail/27980/chicken.
"Classic medieval shield". 2012. *Open Clipart*. https://openclipart.org/detail/169239/classic-medieval-shield.
"Corn". 2011. *Open Clipart*. https://openclipart.org/detail/148705/corn.
"Eagle PNG Clip Art". n.d. *ClipartPNG*. https://clipartpng.com/?1454,eagle-png-clip-art.
"Ear colored". *Open Clipart*. https://openclipart.org/detail/204863/ear-colored.
"Fan-2d". 2014. *Open Clipart*. https://openclipart.org/detail/191035/fan2d.
"Fence clipart #5". n.d. *Design Looter*. http://designlooter.com/fence-clipart.html#gal_post_1363_fence-clipart-7.png.
"Ficedula_hodgsonii" (bird). n.d. *1001 Free Downloads*. https://www.1001freedownloads.com/free-clipart/ficedula_hodgsonii.
"Fox - coloured". 2015. *Open Clipart*. https://openclipart.org/detail/222948/fox-coloured.
"Garlic". 2009. *Open Clipart*. https://openclipart.org/detail/26145/garlic.
"Goat - coloured". 2015. *Open Clipart*. https://openclipart.org/detail/219819/goat-coloured.
"Gravesite in spring". 2015. *Open Clipart*. https://openclipart.org/detail/217615/gravesite-in-spring.
"Grey Canoe Clipart Design". 2013. *Sweet Clip Art*. http://sweetclipart.com/grey-canoe-clipart-design-1438.
"Greyshaded key". 2011. *Open Clipart*. https://openclipart.org/detail/148987/greyshaded-key.
"Hammer and nails". 2008. *Open Clipart*. https://openclipart.org/detail/20591/hammer-and-nails (nails were removed).
"King of hearts". 2008. *Open Clipart*. https://openclipart.org/detail/20311/king-of-hearts (background removed).
"Leaf Cliparts #159482". n.d. *Clipart Library*. http://clipart-library.com/clipart/3574.htm.
"Military style tent". 2017. *Open Clipart*. https://openclipart.org/detail/284413/military-style-tent.
"My guitar". 2010. *Open Clipart*. https://openclipart.org/detail/65095/my-guitar.
"Nose". 2013. *Open Clipart*. https://openclipart.org/detail/184612/nose (color added).
"Pen remix". 2014. *Open Clipart*. https://openclipart.org/detail/204996/pen-remix.
"Pink cake". 2009. *Open Clipart*. https://openclipart.org/detail/23966/pink-cake.
"Pizza Cliparts #212987". n.d. *Clipart Library*. http://clipart-library.com/clipart/436975.htm.
"Scarf - coloured". 2015. *Open Clipart*. https://openclipart.org/detail/214331/scarf-coloured.
"Sailing ship". 2016. *Open Clipart*. https://openclipart.org/detail/248987/sailing-ship-16 (color partly added).
"Snake - coloured". 2015. *Open Clipart*. https://openclipart.org/detail/214285/snake-coloured.
"Sofa-tandem". 2010. *Open clipart*. https://openclipart.org/detail/29845/sofatandem.
"Toad". 2009. *Open Clipart*. https://openclipart.org/detail/22097/toad.
"Cartoon turtle 1". 2016. *Open Clipart*. https://openclipart.org/detail/263073/cartoon-turtle-1.
"Tire". 2010. *Open Clipart*. https://openclipart.org/detail/38071/tire.
"Tree". n.d. *1001 Free Downloads*. https://www.1001freedownloads.com/free-clipart/tree-23.
"Wolf Clipart - Image #11236". n.d. *Cliparting*. http://cliparting.com/free-wolf-clipart-11236.
"Wooden chair". 2012. *Open Clipart*. https://openclipart.org/detail/167460/wooden-chair.
"Young lamb". 2006. *Open Clipart*. https://openclipart.org/detail/1911/young-lamb.

11. Appendix

Item	paired with	Item	paired with
sircle		sunset	
berd		ball	
rane		river	
brane		bread	
boan		beer	
chare		chain	
noze		mask	
soope		skull	
focks		fence	
fone		face	

Item	paired with	Item	paired with
hownd		hedge	
phish		fruit	
bair		book	
tode		tyre	
kee		cup	
angker		orange	
aks		egg	
kanoo		camel	
gittar		garlic	
ere		owl	

Item	paired with	Item	paired with
spunj		spear	
sord		sofa	
leef		lamb	
baik		belt	
brume		bench	

I made great efforts to select only images that may be copied for non-commercial use. However, images will be removed immediately if a copyright infringement is reported.

Abstract

The process of reading has been researched quite extensively over the past decades. It is a skill that evolved fairly late in human history, but over the past centuries, it has become a skill that people use on a daily basis for a variety of purposes. In investigating how a reader processes unfamiliar words, the reading of nonwords is commonly compared to the reading of pseudohomophones that sound like real words, but are spelled differently (e.g., *brane*). Primarily, the objective is to observe how unfamiliar letter combinations are translated into sounds, and if there is a difference between the reading of words that possess a known phonological pattern and words that do not. Theoretically, this known phonological pattern would allow a reader to activate semantic information as well, but the aspect of meaning-making is typically neglected in studies of reading aloud. Semantics is assumed to be a byproduct of reading, and its role in models of reading aloud is seen as insignificant. As a result, few efforts have been made to develop an integrative approach to reading aloud that includes orthography, phonology, and semantics. This thesis therefore puts an emphasis on semantics in the area of reading aloud unfamiliar words. It presents three models – two dual-route models (DRC, CDP+) and one integrated model (PDP) – and investigates the role of semantics in these models. Moreover, the so-called pseudohomophone effect is explained, and how the models account for this effect. To shed light on how meaning is activated during reading, the organization of the mental lexicon is explored.

The experimental part of this thesis combines a naming task with a picture task; real words and pseudohomophones are used as stimuli in a mixed-list design. In the naming task, no effect of (base)word frequency was found for either word type, which suggests that such an effect simply cannot appear in a mixed-list design (but only in research designs in which all words of the same type are presented in one block). Moreover, it was found that participants remembered pseudohomophones twice as frequently as real words, even though they had not received any instructions to memorize words. This surprising result is attributed to the increased cognitive effort involved in the processing of pseudohomophones. In the picture task, participants showed little difficulty in providing the correct answer, not even when pseudohomophones were followed by a picture of a semantically related object (e.g., *sord* / shield). Taking all this evidence into account, it is concluded that meaning is activated to a large extent in the reading of pseudohomophones. This conclusion is illustrated by means of the 'petal metaphor': the essence of a word is composed of various aspects, such as orthography, phonology, semantic features, sensory information, connotation, or personal associations. All of these aspects are

interconnected – as if they were petals of the same flower – and therefore, a wholesome model of reading aloud that links these aspects ought to be developed.

Deutsche Zusammenfassung (German abstract)

Der Prozess des Lesens wurde während der vergangenen Jahrzehnte intensiv untersucht. Das Lesen ist eine Fähigkeit, die sich erst spät in der menschlichen Geschichte entwickelt hat, sich aber während der vergangenen Jahrhunderte etabliert hat und von Menschen täglich für eine Vielzahl von Zwecken eingesetzt wird. Um zu untersuchen, wie ein Leser unbekannte Wörter verarbeitet, wird das Lesen von Nichtwörtern üblicherwerweise mit dem Lesen von Pseudohomophonen, die wie echte Wörter klingen, aber anders geschrieben werden (z.B. *brane*), verglichen. Im Wesentlichen wird dabei untersucht, wie unbekannte Buchstabenkombinationen in Laute übersetzt werden, und ob ein Unterschied im Lesen von Wörtern mit bekanntem phonologischem Muster und Wörtern ohne ein solches Muster besteht. Theoretisch könnte dieses bekannte phonologische Muster es einem Leser erlauben, auch semantische Informationen zu aktivieren, aber der Aspekt des Verstehens wird normalerweise in Studien über lautes Lesen vernachlässigt. Es wird angenommen, dass die Semantik ein Nebenprodukt des Lesens ist, und ihre Rolle in Lesemodellen wird als unbedeutend angesehen. Aus diesem Grund wurden wenige Anstrengungen unternommen, einen intergrativen Ansatz zum lauten Lesen zu entwickeln, der Orthographie, Phonologie und Semantik beinhaltet. Diese Masterarbeit setzt daher einen Schwerpunkt auf Semantik im Bereich des Vorlesens unbekannter Wörter. Sie präsentiert drei Modelle – zwei Modelle mit dualer Route (DRC, CDP+) und ein integriertes Modell (PDP) – und untersucht die Rolle der Semantik in diesen Modellen. Darüber hinaus wird der sogenannte Pseudohomophon-Effekt erläutert, und wie die Modelle diesen Effekt erklären. Auch die Organisation des mentalen Lexikons wird genauer beleuchtet, um zu untersuchen, wie Bedeutung während des Lesens erfasst wird.

Der experimentelle Teil dieser Masterarbeit verbindet eine Vorlese-Aufgabe mit einer Bilderaufgabe; echte Wörter und Pseudohomophone werden als Stimuli in einem Forschungsdesign mit gemischter Liste eingesetzt. In der Vorlese-Aufgabe wurde für beide Wortarten kein Effekt von (Basis-)Worthäufigkeit gefunden, was nahelegt, dass ein solcher Effekt schlichtweg nicht in einem Forschungsdesign mit gemischter Liste auftreten kann (sondern nur in Designs, bei denen alle Wörter des selben Typs in einem einheitlichen Block präsentiert werden). Des Weiteren wurde festgestellt, dass Teilnehmer sich doppelt so oft an Pseudohomophone erinnerten wie an echte Wörter, obwohl sie keinerlei Instruktionen erhalten

hatten, sich Wörter zu merken. Dieses überraschende Ergebnis wird dem höheren kognitiven Aufwand zugeschrieben, den die Verarbeitung von Pseudohomophonen mit sich bringt. In der Bilderaufgabe hatten die Teilnehmer wenig Mühe, die korrekte Antwort zu geben, nicht einmal wenn einem Pseudohomophon das Bild eines semantisch ähnlichen Objekts folgte (z.B. *sord* / shield). Die Berücksichtigung aller Erkenntnisse legt die Schlussfolgerung nahe, dass Bedeutungaspekte beim Lesen von Pseudohomophonen in hohem Maße aktiviert werden. Diese Schlussfolgerung wird mit der ‚Blütenblätter-Metapher' verdeutlicht: die Essenz eines Wortes setzt sich aus vielerlei Aspekten zusammen, wie etwa Schriftbild, Phonologie, semantische Merkmale, sensorische Informationen, Konnotation, oder persönliche Assoziationen. All diese Aspekte sind miteinander verbunden – als wären sie Blüttenblätter der selben Pflanze – und daher sollte für das laute Lesen ein ganzheitliches Modell entwickelt werden, das diese Aspekte miteinander verbindet.